ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

VOLUME 88, ART. 3 PAGES 533-770

Managing Editor
FRANKLIN N. FURNESS

Associate Editor
THOMAS SIMPSON

AMINO ACIDS, PEPTIDES, AND PROTEINS

LIST OF AUTHORS

KARL FOLKERS (*Conference Chairman*), V. G. ALLFREY, G. W. ANDERSON, J. APGAR, M. BODANSZKY, R. D. COLE, L. C. CRAIG, B. P. DOCTOR, P. EDMAN, J. H. FRENSTER, M. GOODMAN, C. H. W. HIRS, K. HOFMANN, R. W. HOLLEY, J. W. HOPKINS, H. A. ITANO, T. P. KING, W. KONIGSBERG, L. MENDIOLA, A. E. MIRSKY, D. NATHANS, R. RENDI, R. B. ROBERTS, E. ROBINSON, E. E. SCHMITT, J. C. SHEEHAN, M. L. STEPHENSON, V. DU VIGNEAUD, R. C. WARNER, D. A. YPHANTIS, AND P. C. ZAMECNIK.

Consulting Editor
KARL FOLKERS

NEW YORK

PUBLISHED BY THE ACADEMY

August 31, 1960

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

VOLUME 88, ART. 3 PAGES 533–770

August 31, 1960

Managing Editor

FRANKLIN N. FURNESS

Associate Editor

THOMAS SIMPSON

AMINO ACIDS, PEPTIDES, AND PROTEINS*

Consulting Editor and Conference Chairman

KARL FOLKERS

CONTENTS

* This series of papers is the result of a conference on *Amino Acids, Peptides, and Proteins* held by The New York Academy of Sciences on February 25 and 26, 1960.

Part IV. Biosynthesis of Proteins

INTRODUCTORY REMARKS

Karl Folkers
Merck, Sharp and Dohme Research Laboratories, Rahway, N. J.

Symposia and publications on amino acids and their complex combinations are by no means uncommon, and their frequency may be increasing. As in many fields of natural products, when one thinks of amino acids and their combinations, one may be concerned with their isolation and purity, or with their degradation and structure, or with their synthesis. The subdisciplines of isolation, structure, and synthesis comprise much of the essential chemistry of these natural products, and they provide a basis for the future biological and medical study of these acids.

A monograph embracing new data and trends on amino acids and their combinations has seemed overdue, particularly one including such information on the peptides and proteins based on the three subdisciplines: isolation, structure, and synthesis.

It is of course more common for publications to include contributions covering a single discipline or subdiscipline. For this monograph it seemed highly desirable to bring together contributions from separate areas of the broad field of amino acids, peptides, and proteins, so that more investigators in one subdiscipline might obtain new ideas and stimulation from workers in other fields. We all know that unexpected progress is made when investigators from different areas stimulate each other.

The first three session chairmen, Bell, Brink, and John C. Sheehan, have organized timely presentations. More and more frequently the chemist on natural products associates the knowledge of organic syntheses and of biosyntheses, and it seemed that for this publication a fourth section on the biosynthesis of proteins would be indispensable. A valuable section on this subject was accordingly outlined by Lipmann.

For the sake of brevity, may I comment on only two of the papers in this monograph. The importance of homogeneity in the chemistry considered in this monograph is such that a careful consideration of it should be a formal part of its content. Continued awareness of this aspect of natural products is encouraged. In fact, the importance of homogeneity is not limited to natural materials alone, but is also significant for the chemist synthesizing peptides, particularly when he achieves a multicomponent product. The name of Lyman Craig is so firmly linked to aspects of homogeneity that Bell and I considered that no one could bring us up to date better in this aspect than Craig himself.

The late James Sumner isolated the enzyme urease some thirty-odd years ago, and biochemists in the few years following this significant event witnessed the polemical reception of Sumner's interpretation of the protein nature of this enzyme. After not too long a period, biochemists are now witnessing the structural elucidation of enzymes, and readers of this publication may take particular note of a paper that details the complete covalent structure of the first enzyme having a known substrate to be elucidated. An account of the

124 amino acid-sequence of ribonuclease by C. H. W. Hirs is also included. This sequence is an outstanding accomplishment, and we now may expect biochemists to undertake structural peptide problems of even greater complexity.

A few years ago, it was written: "Research workers are usually so busy piling brick on brick on the edifice of human knowledge that there is never time for them to stand back and survey what has been built and how it has been done."

For this monograph, it was considered desirable to have an initial paper by someone who could survey what the biochemists working on amino acids, peptides, and proteins have built and how they have done it: someone with the experience of personal contributions and intellectual insight in this field might help us appreciate the over-all perspective as we pile brick on brick, or condense amino acid on amino acid. Vincent du Vigneaud kindly consented to keynote this publication. That du Vigneaud is uniquely qualified to do so is readily evident when one recalls his extensive contributions in all the areas of isolation, structure, organic synthesis, and biosynthesis, including research on many individual amino acids, glutathione, insulin, transmethylation, penicillin, and biotin. History will note particularly his contributions to the posterior pituitary hormones, recognized by a Nobel Prize in 1955, "especially for the first synthesis of a polypeptide hormone."

This contributor's academic training was received at universities and institutes in the United States and Germany. His research has been recognized also by the Mead Johnson, Borden, Lasker, Osborne and Mendel, Scott, Passano, Nichols, and Gibbs medals and awards. He has received honorary doctorates and has been active in various scientific societies.

EXPERIENCES IN THE POLYPEPTIDE FIELD: INSULIN TO OXYTOCIN

Vincent du Vigneaud
Cornell University Medical College, New York, N. Y.

It has been my privilege for the past 35 years to have witnessed, and at times participated in, a truly exciting development in the field of biochemistry. The development I have in mind involved: first, the recognition and acceptance that a hormone could be a protein or polypeptide; second, the working out of the organic chemical structures of certain polypeptide and protein hormones; third, the demonstration by synthesis of the authenticity of the structures postulated for certain lower molecular weight polypeptide hormones; and, finally, the present stage where the structures of these polypeptides can be modified almost at will by synthetic means in an effort to understand the specificity of the structures involved—by the replacement of constituent amino acids by other amino acids to produce various analogues and derivatives.

Parallel to this development of our knowledge of the structure of the polypeptide hormones has been a comparable development, the recognition that enzymes are proteins, culminating in the recent achievement of the establishment of the structure of one of these, ribonuclease. Closely following the unfolding of our knowledge of hormones and enzymes have been the exciting developments in the field of viruses. We are even beginning to get a glimpse of the interrelationship between a protein and a nucleic acid in their role in virus activity. Before long we shall know the structures of certain virus proteins. Already that is well on the way. Synthesis may well follow. These are truly exciting times. It is appropriate that we consider in this symposium certain phases of this rapidly expanding polypeptide field. Perhaps there is some advantage in pausing initially to reflect on where we have been, to ponder on where we are, and to speculate on where we are going. Karl Folkers has asked me to do this and has particularly urged me to discuss from personal experience the development of the protein and polypeptide hormone field as I have lived through it.

As one surveys the protein and polypeptide hormone field today, it is almost unbelievable how hard it was some 35 years ago for the workers in the field to accept the idea that a protein could be a hormone or, to put it the other way around, that a hormone could be a protein. Today we have arrived at the point where we have several polypeptide hormones already synthesized, and we know the structures of several other polypeptide and protein hormones and even one enzyme. There is little doubt that in the not too far distant future the protein hormones and enzymes whose structures are established by degradative studies will succumb to the synthetic attack of the peptide chemist.

It is, of course, a tradition of organic chemistry that before final acceptance of any structure, synthesis by unambiguous procedures must be accomplished. It is reassuring to realize that the organic chemistry of peptides is now reaching the stage where synthetic proof of structure can be called upon to check the

peptide structures arrived at through degradative means and heretofore considered too complex to be subjected to the synthetic test.

Insulin played a key role in forcing the acceptance of the fact that a protein could be a hormone. The history of insulin was a tortuous one from the time of the recognition by von Mering and Minkowski[1] in 1889 that the pancreas produced some substance necessary for the proper metabolism of glucose. These investigators and many others tried to obtain active extracts but to no avail. The subsequent literature reveals attempt after attempt to accomplish this, but failure after failure resulted. However, there were enough encouraging signs to maintain hope of obtaining active extracts until the crowning achievement of Banting and Best[2] in Toronto in 1921 and of Murlin at about the same time, independently, in Rochester, N. Y., as described fully in a later review.[3] Throughout the intervening years from 1889 to 1921 it had begun to be suspected that the proteolytic enzymes present in the acinous tissue of the pancreas were destroying the active principle. It was the avoidance of this destruction that led to the successful extraction of what was called insulin.

Then came the race to isolate the pure hormone from these extracts. Scientists all over the world sought to accomplish this. The first to achieve success was Abel[4] at The Johns Hopkins University Medical School, Baltimore, Md., in 1926 when he isolated insulin in crystalline form. It is curious that with all the early evidence that this elusive substance could be destroyed by proteolytic enzymes so many workers were reluctant to believe that Abel had actually isolated insulin, because it appeared to be a protein. Many thought that he had isolated a protein with an unknown still more highly active hypoglycemic substance adsorbed on the surface of the crystals. This same doubt was cast on James Sumner's work with crystalline urease. Sumner was subjected for a number of years to criticism bordering on ridicule because he claimed that his crystals of a protein actually represented crystals of the enzyme urease.

When Abel's announcement of the crystallization of insulin appeared in 1926, I was working at the University of Rochester Medical School in Murlin's laboratory in attempts to isolate insulin and to find out something of its chemical nature. I had isolated fairly potent preparations of insulin and was directing my attention particularly to the nature of the sulfur at various stages of purification. As I purified insulin, the cystine content increased, as determined by the Sullivan method,[5] and I proceeded on the assumption that insulin was a cystine compound. Before Abel's announcement of the crystallization, Abel and Geiling[6] had published a paper in which they were convinced that insulin was a strange sulfur compound because its sulfur, unlike that of cystine, was quite labile, being split out readily as sulfide with dilute sodium carbonate. In this paper they even suggested that perhaps diabetes mellitus was due to the absence of a particular labile sulfur compound in the diet. However, Brand and Sandberg[7] showed that peptides of cystine were more labile toward alkali than cystine; therefore lability of the sulfur did not rule out the presence of cystine peptides. I found that when insulin was heated with dilute alkali the cystine present in the material disappeared, and when insulin was hydrolyzed with acid the sulfur of insulin became stable, as in cystine. I concluded

that insulin was a cystine derivative with cystine linked through peptide linkage to the rest of the molecule.[8] At this stage I can remember what a thrill it was to read Abel's announcement of the isolation of insulin in crystalline form. I tried his brucine-pyridine acetate method, succeeded in confirming his crystallization,[8] and found that the crystals contained a fairly large amount of cystine in bound form as determined by the Sullivan colorimetric method.

In the fall of 1927, it was my privilege to work in Abel's laboratory and there to isolate crystalline cystine from a hydrolysate of crystalline insulin, thus showing beyond doubt that insulin was truly a cystine-containing compound.

In this work during 1927–1928 at The Johns Hopkins Medical School, I collaborated with Oscar Wintersteiner and Hans Jensen.[9-11] This work led not only to the isolation of cystine and tyrosine from crystalline insulin but also to the isolation of histidine, leucine, lysine, and arginine and a study of the nitrogen distribution of insulin by the Van Slyke procedure. The hormone gave a distribution quite in keeping with that of a protein. There was nothing very unusual about insulin from this standpoint.

As we went our separate ways, we all continued to search for a number of years for the presence of some constituent in insulin other than ordinary amino acids, but to no avail. Other investigators joined the search but with like results. It gradually became clear that insulin was made up of residues of ordinary amino acids and ammonia and that it must be the arrangement of the amino acids within the structure of the molecule that conferred the hormonal properties on it.

As time went on I focused attention more and more on accounting for the sulfur of insulin, which proved to be no easy task. It took about 10 years of fairly intensive work to establish that all of the sulfur could be accounted for by cystine, as we reported in 1939 with Miller and Rodden.[12] It is difficult to realize the amount of effort involved with the methods then available. How easy it would be now with the newer tools!

Along with this work through the 1930's and early 1940's, the full amino acid composition was worked out, accounting quantitatively for the molecule, chiefly through intensive efforts in the United States by Brand,[13] in England by Chibnall,[14] and in France by Fromageot,[15] and their respective co-workers. Determination of the actual structure seemed out of the question. Then came the brilliant and classic work in Chibnall's laboratory by Sanger and his co-workers.[16] Through their use of the dinitrophenyl method for end-group analysis, they were able to establish that insulin was made up of only two different kinds of chains: one with glycine N-terminal and the other with phenylalanine N-terminal. Jensen and Evans[17] had earlier demonstrated that a phenylalanine radical with a free amino group was present in insulin but, with the techniques then available, could not ascertain that there was present only one other kind of end α-amino group. Sanger and his co-workers later separated the two types of chains through oxidation and worked out the sequence of the amino acids within each.[16] The so-called A chain of 21 amino acid residues contained four cysteic acid residues and the B chain of 30 amino acid residues, two cysteic acid residues. How then were the chains linked together? The molecular weight at that time was regarded as 12,000. I must say that this

daringness of Sanger is the quality that most astonished me: that he dared to attack the structure of a molecule with a molecular weight of 12,000. Earlier workers had been confronted by a value for the molecular weight of 36,000 to 48,000, and the prospect of ever establishing sequence then seemed quite remote. Establishment of the fact by Harfenist and Craig[18] in 1952 that insulin could exist under certain conditions as a molecule with a molecular weight of 6000 enabled Sanger and his co-workers to arrive at his now well-known structure of two chains, one of 20 amino acids and the other of 30 amino acids, linked through two disulfide bridges with the chain of 20 amino acids containing an internal disulfide bridge. Otherwise, when Sanger and his group had finally established the disulfide bridges, after surmounting the complication of sulfur interchange, they would have had to put four chains together to arrive at a structure satisfying the molecular weight of 12,000, and this would have led to quite a different structure.

The next challenge with regard to insulin is obviously its synthesis. One has a right to dare to think that insulin can be synthesized, even though in comparison with the synthesis of oxytocin the difficulty appears to be of another order of magnitude. The chains themselves should not be too difficult to make, but the real question is the bringing about of the linkages involved in the three disulfide bridges. Already one can discern that several laboratories are picking up the gauntlet, and I am sure it is only a matter of time before synthetic insulin will be a reality.

It is through my interest in insulin that I became interested in the posterior pituitary hormones. This was in 1932. We had found a bit earlier that the activity of insulin was destroyed by treatment of the hormone with cysteine or glutathione at room temperature and at a *p*H close to neutral and, furthermore, that the activity was not restored on reoxidation by aeration.[19] I then wondered if this would be of general significance for protein or polypeptide hormones. I knew the background of the posterior pituitary hormones from my stay with Abel and was aware of the contribution of Kamm and his co-workers[20] in obtaining one fraction of highly potent oxytocic activity with very little pressor activity and another fraction highly potent from the pressor standpoint but having very little oxytocic activity.

The evidence indicated that they might be polypeptides, and I asked Kamm if they contained cystine. Kamm told me that his preparations contained a fairly high percentage of sulfur but that cystine had not been detected in them. I asked him then if he would be willing to let me have some of the purified fractions to determine, from the experience we had had with cystine in insulin, whether they contained cystine. Kamm let me have these preparations, and we were able to show that cystine accounted for most of the sulfur in both.[21,22] We then treated the oxytocic and pressor fractions with cysteine under the conditions we had treated insulin. We were not able to detect much change, if any, under these conditions in their respective activities.[23] This then aroused our interest in isolating these hormones to find out if they truly contained cystine linked only to other amino acids, as appeared to be the case with insulin. The posterior pituitary hormones were particularly attractive compounds since the evidence indicated that they were far smaller

in molecular size than insulin. Kamm[24] had carried out experiments comparing the rate of diffusion of the hormones to that of epinephrine and had concluded that their molecular weight was in the neighborhood of 600. Thus, if they could be isolated and their structures worked out, perhaps these smaller molecules might lend themselves to a synthetic attack, and we could make more headway with these compounds in a study of the relation of structure or architecture of amino acid combinations to biological activity than we could with insulin, which appeared to be so much more complex. We therefore attacked the isolation of the posterior pituitary hormones initiating a study of their chemical and physical behavior with the encouragement of both Abel and Kamm, two men who had pioneered in this field.

In the pre-World War II years we made considerable progress with purification of these hormones by electrophoretic techniques.[22,25-27] After the war, we turned again to their isolation and, utilizing Craig's countercurrent distribution,[28] we succeeded in isolating oxytocin in collaboration with Livermore[29] and vasopressin with the cooperation of Turner and Pierce.[30]

Having achieved isolation of these hormones, we were able to determine their individual biological effects, and we found that oxytocin is not only the principal uterine-contracting hormone but also the chief milk-ejecting one. This hormone is also the main avian-depressor principle of the posterior pituitary. We found vasopressin to be the principal pressor-antidiuretic hormone; that is, the one that is the most powerful in increasing the blood pressure and, what is extremely important, the one that is involved in kidney function. Each hormone has qualitatively the activities of the other, but quantitatively the activities are present in the two hormones to quite different extents.

Having purified the hormones, we turned to the study of their composition. We were able to establish the fact that oxytocin[31] and vasopressin[30] were both made up of eight amino acids in equimolar ratios to each other and ammonia in a molar ratio of three to any one amino acid. The elegant starch-column chromatographic procedures for quantitative amino acid analysis on milligram quantities of Moore and Stein made this possible.

We found that six of the eight amino acids were identical in both hormones but that, in vasopressin, phenylalanine and arginine appeared in place of isoleucine and leucine, which were present in the oxytocin. This was the case in vasopressin isolated from beef glands. In hormones isolated from hog glands, we found, with Popenoe and Lawler,[32] that the vasopressin contained lysine in place of arginine, an interesting species difference; with Pierce and Gordon[33] we found that the oxytocin from the hog glands was the same as that from beef. More recently we have shown, with Light,[34] that vasopressin isolated from human pituitaries is arginine-vasopressin.

Knowing the amino acid composition of hydrolysates of oxytocin and vasopressin, we naturally wanted to know how these constituent amino acids and ammonia were linked together in the original hormonal molecule. Obviously the molecular weight was as crucial here as it had been in insulin. Paul Bell* was kind enough to determine a molecular weight of oxytocin for us, and his result was later confirmed by Rachele in our laboratory. The value agreed

*. American Cyanamid Laboratories, Stamford, Conn.

with a molecule containing one of each of the eight amino acid residues and three ammonia residues.

The structures of these hormones were elucidated only through several years of intensive study with various collaborators, in the early stages with Pierce, Mueller, Davoll, and Turner,[35-38] in the final decisive stages of the oxytocin work with Ressler and Trippett,[39,40] and on the vasopressin with Popenoe and Lawler.[41-44]

The degradative studies on the polypeptide included oxidation with performic acid, desulfurization with Raney nickel, determination of terminal groups by the dinitrophenyl method, degradation with bromine water, and determination of sequence of amino acids by Edman degradation and by partial hydrolysis with acid.

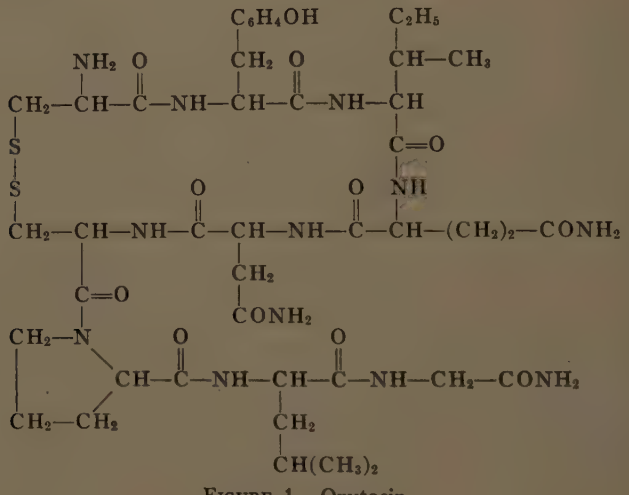


FIGURE 1. Oxytocin.

The structure of oxytocin which we arrived at by our degradative studies on the hormone is shown in FIGURE 1.

To arrive at this structure, we had to assume: (1) that we had not missed some small fragment that did not give a ninhydrin reaction like an acetyl or formyl grouping; (2) that the two ammonia residues unaccounted for degradatively were attached to the glutamic and aspartic acid residues and that glutamine and asparagine were present and not isoglutamine or isoasparagine; (3) that each amino acid was connected to another through ordinary peptide bonds and, specifically, that the bromine water cleavage of the bond between tyrosine and isoleucine did not signify some unusual nature of the peptide bond; and (4) that no rearrangement had occurred in the oxidation with performic acid, in the desulfurization, or in the partial hydrolysis of oxytocin, oxidized oxytocin, and desthio-oxytocin. With these assumptions, this 20-membered cyclic structure, not previously encountered in nature, was the only structure that explained the data. Nevertheless it was these assumptions that made us feel that synthesis was mandatory for proof of structure.

The structure we postulated for arginine-vasopressin is shown in FIGURE 2. It is apparent how closely the postulated structures of these two hormones resemble one another. The structure we suggested for lysine-vasopressin, the hormone from hog posterior pituitaries, was the same as that for arginine-vasopressin, except that lysine appeared in the side chain in place of arginine.[44]

The same structure for oxytocin was proposed independently by Tuppy[45] through a similar rationalization of degradative data. Tuppy's proposal was based on the data from our laboratory on composition, molecular weight, and terminal groups, and on the cyclic structure involving the disulfide linkage, together with his independent studies on the sequence of amino acids involving partial hydrolysis with acid and with an enzyme. The interpretation of the data and the assumptions made were parallel in the two laboratories. The

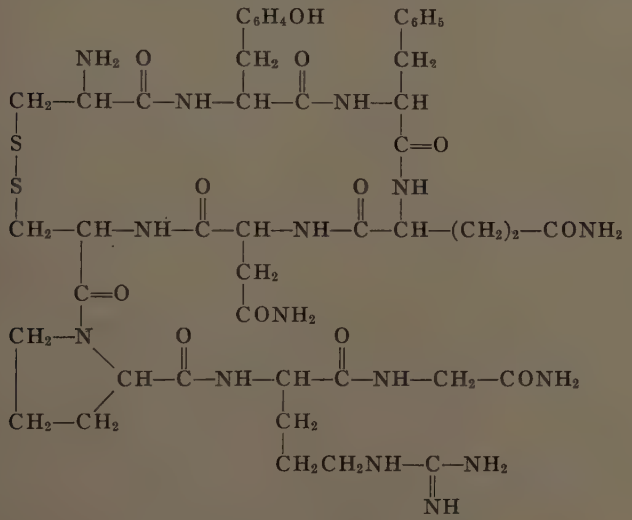


FIGURE 2. Arginine-vasopressin.

same structure for arginine-vasopressin involving like assumptions was also suggested independently by Acher and Chauvet[46] in Fromageot's laboratory in Paris.

Our clue to the synthesis of oxytocin rested on the observation we had made with Sealock[23] in the 1930's that oxytocin could be reduced and that upon reoxidation the biological activity remained.

On the basis of our postulated structure for oxytocin, the reduction and reoxidation of the hormone could be interpreted as involving the opening and closing of the twenty-membered ring containing the disulfide linkage. On reduction the octapeptide becomes a nonapeptide, since the amino acid cystine on reduction gives rise to 2 molecules of cysteine, simply a point of nomenclature.

Taking advantage of the clue that reduced oxytocin was undoubtedly convertible to oxytocin itself, we felt that we could approach the synthesis of oxytocin through the reduced form, but we also knew that if we were to accomplish this we would have to protect the sulfhydryl groups of both cysteine

residues and the free amino group of one of the cysteine residues throughout the various synthetic steps. Our own earlier work with Sifferd[47] in 1935 had shown that a carbobenzoxy group could be removed from an amino group with sodium in liquid ammonia and that benzyl groups could be removed from

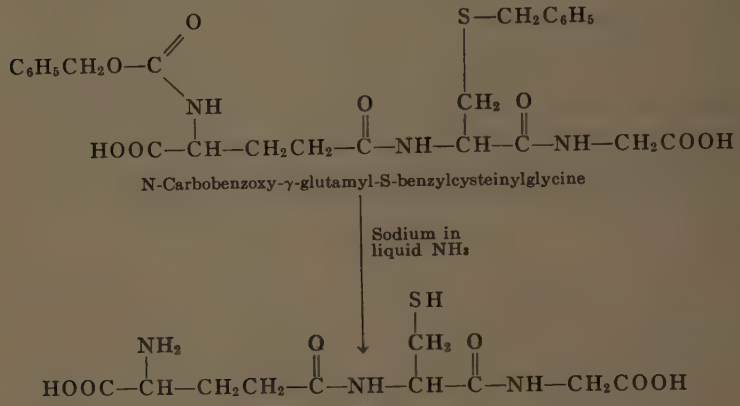


FIGURE 3. Glutathione.

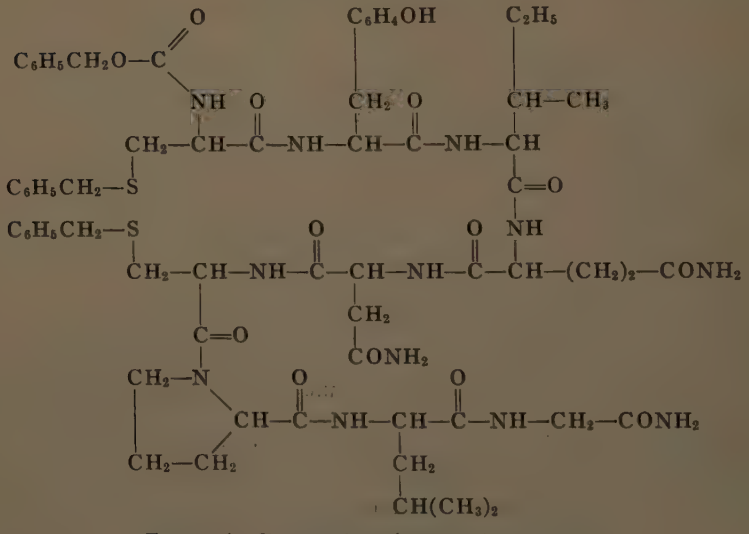


FIGURE 4. Protected nonapeptide intermediate.

benzylthio ethers by the same treatment. In fact, we succeeded in synthesizing glutathione with Miller[48] in 1936 by the employment of such protecting groups. Our protected intermediate in that case was carbobenzoxy-γ-L-glutamyl-S-benzyl-L-cysteinylglycine. Reduction of this compound by sodium in liquid ammonia gave rise to glutathione, as shown in sequence of reactions shown in FIGURE 3.

Therefore we felt that if we could synthesize the linear protected nonapeptide (FIGURE 4) with benzyl groups protecting the cysteine sulfurs and a carbo-

benzoxy group on the amino groups of one of the cysteine residues, we ought to be able to convert that intermediate to the sulfhydryl form of the hormone by reduction with sodium in liquid ammonia and then oxidize the reduced form to the disulfide form, which would be oxytocin.

In work with Gordon,[49] reduced oxytocin (which we now find it convenient to call oxytoceine) was benzylated and the dibenzyl derivative was isolated. This compound was biologically inactive. We then demonstrated that the derivative could be reconverted to oxytocin by sodium in liquid ammonia followed by oxidation.

I shall not go into detail as to how this nonapeptide was put together, but I shall recall to you the general method of approach used in the synthetic studies with Ressler, Swan, Katsoyannis, and Roberts.[50-54] The tetrapeptide amide S-benzyl-L-cysteinyl-L-prolyl-L-leucylglycinamide was condensed with tosyl-L-isoleucyl-L-glutaminyl-L-asparagine to give the protected heptapeptide amide. The free S-benzyl heptapeptide amide obtained from the latter was condensed with *N*-carbobenzoxy-S-benzyl-L-cysteinyl-L-tyrosine to give our desired protected nonapeptide *N*-carbobenzoxy-S-benzyl-L-cysteinyl-L-tyrosyl-L-isoleucyl-L-glutaminyl-L-asparaginyl-S-benzyl-L-cysteinyl-L-prolyl-L-leucylglycinamide. If our thinking was correct, then, when this protected nonapeptide was dissolved in liquid ammonia and reduced with metallic sodium we should get reduced oxytocin; then, upon aeration of an aqueous solution of the reduced oxytocin we should get oxytocin itself. This reduction and subsequent oxidation was carried out, and I can tell you that we were delighted to find that the solution was biologically active. Next came the hard job of getting the synthetic material out of the reaction mixture. This was finally accomplished, and we were able to demonstrate by a battery of tests—chemical, physical, and biological—that it was identical to the natural oxytocin.

We therefore felt justified in our conclusions that the synthetic octapeptide amide was oxytocin and that the structure we believed it to possess was truly the correct one.

Since this original synthesis of oxytocin, the hormone has been synthesized in our own laboratory and in other laboratories by various modifications. For example, with Bodanszky[55] we worked out an improved synthesis condensing a protected carbobenzoxy pentapeptide with the original tetrapeptide patterned somewhat after the 5 + 4 synthesis of lysine vasopressin we had worked out with Bartlett, Jöhl, Roeske, Stedman, Stewart and Ward.[56,57] More recently we have presented with Bodanszky[58,59] a stepwise synthesis from the glycine end adding one amino acid at a time utilizing the nitrophenyl ester method, a method of peptide synthesis originally introduced by Bodanszky[60] in Hungary. I believe that Anderson, then at Stamford, Conn., now at Pearl River, N. Y., was the first to confirm (in unpublished work) our synthesis of oxytocin from the protected nonapeptide, producing the latter by a slightly different procedure condensing, as I recall it, the protected tripeptide with a hexapeptide. Syntheses have also been accomplished by Boissonnas[61] in Switzerland, Rudinger[62] in Czechoslovakia, Bodanszky[63] in Hungary before he came to Cornell University Medical College, Velluz[64] in France, and Beyerman[65] in the Netherlands, with their respective colleagues. Essentially, the same nonapeptide was synthesized, differing in some cases in the protecting groups

employed and in the intermediates used. After removal of the protecting groups, the last step involving conversion by oxidation to the octapeptide oxytocin was carried out in all the methods according to the method developed in our laboratory with Gordon for the regeneration of oxytocin from the dibenzyl derivative of the reduced natural oxytocin.

We have also synthesized the vasopressins, both the arginine and the lysine, by methods somewhat similar but more complicated because of the basic amino acids involved, and the isolated synthetic compounds were identical with their corresponding natural hormones.[56,57,66]

The synthesis of oxytocin and the vasopressins has given final proof beyond question that ordinary amino acids, when hooked together by the common peptide linkage in a certain pattern, give rise to polypeptides of remarkable pharmacological properties. Until synthesis of one of the polypeptide or protein hormones was achieved, there could always lurk the suspicion that perhaps some strange linkage was involved not yet uncovered by degradation studies or that some small fragment vital to the special activity had escaped detection. This was more real than might be casually apparent. Furthermore, synthesis of a proteinlike substance according to the structure arrived at by degradative studies has given greater credence to the structures determined by similar approaches for other proteins but not yet confirmed by synthesis.

The synthesis of oxytocin brought the hormone field to the stage where the fascinating problem of the specificity of structure in relation to pharmacological activity can be approached in an almost unlimited way. As soon as we accomplished the synthesis of oxytocin we undertook the synthesis of related compounds pertinent to this question. Studies in this direction were also undertaken in other laboratories such as those of Biossonnas and his colleagues in Switzerland and Rudinger and his colleagues in Czechoslovakia.*

Synthesis of analogues and homologues will also be of value in providing substrates for the study of the action of certain enzymes. Synthesis of the hormones and related compounds labeled in specific positions in the molecule with tritium, radiocarbon, and radiosulfur will open new avenues of study of the site of action, metabolism, and fate. Synthesis of desired analogues modified in specific ways may be of invaluable help to the biophysicist, particularly to the X-ray diffraction worker. Synthesis of a hormone has, of course, its practical side in making us independent of natural sources which is vital as use becomes more widespread. Furthermore, synthesis of hormone analogues and derivatives may aid us in reaching the tantalizing goal of anti-hormones such as anti-oxytocics or anti-antidiuretics (diuretics) with possible valuable clinical attributes.

We are now at this stage of development in the polypeptide and protein hormone field, and the future undoubtedly holds exciting possibilities.

References

1. von Mering, J. & O. Minkowski. 1889. Centr. klin. Med. **10:** 393.
2. Banting, F. G. & C. H. Best. 1921–1922. J. Lab. Clin. Med. **7:** 251, 764.

* A brief survey of some of the analogues and homologues of oxytocin and vasopressin prepared at Cornell University Medical College and elsewhere was given at this point in the original presentation.

3. Murlin, J. R. 1923. Endocrinology. **7:** 519.
4. Abel, J. J. 1926. Proc. Natl. Acad. Sci. U. S. **12:** 132.
5. Sullivan, M. X. 1926. Public Health Repts., U. S. **xli:** 1030.
6. Abel, J. J. & E. M. K. Geiling. 1925. J. Pharmacol. Exptl. Therap. **25:** 423.
7. Brand, E. & M. Sandberg. 1926. J. Biol. Chem. **70:** 381.
8. du Vigneaud, V. 1927. J. Biol. Chem. **75:** 393.
9. du Vigneaud, V., H. Jensen & O. Wintersteiner. 1928. J. Pharmacol. Exptl. Therap. **32:** 367.
10. Jensen, H., O. Wintersteiner & V. du Vigneaud. 1928. J. Pharmacol. Exptl. Therap. **32:** 387.
11. Wintersteiner, O., V. du Vigneaud & H. Jensen. 1928. J. Pharmacol. Exptl. Therap. **32:** 397.
12. du Vigneaud, V., G. L. Miller & C. J. Rodden. 1939. J. Biol. Chem. **131:** 631.
13. Brand, E. 1946. Ann. N. Y. Acad. Sci. **47**(2): 187.
14. Chibnall, A. C. 1942. Proc. Soc. Roy. London. **B131:** 136.
15. Acher, R., C. Fromageot & M. Jutisz. 1950. Biochim. et Biophys. Acta. **5:** 81.
16. Sanger, F. *In* Les Prix Nobel en 1958. : 134–146. Imprimerie Royale P. A. Norstedt and Söner. Stockholm, Sweden.
17. Jensen, H. & E. A. Evans. 1935. J. Biol. Chem. **108:** 1.
18. Harfenist, E. J. & L. C. Craig. 1952. J. Am. Chem. Soc. **74:** 3087.
19. du Vigneaud, V., A. Fitch, E. Pekarek & W. W. Lockwood. 1931. J. Biol Chem. **94:** 233.
20. Kamm, O., T. B. Aldrich, I. W. Grote, L. W. Rowe & E. P. Bugbee. 1928. J. Am. Chem. Soc. **50:** 573.
21. du Vigneaud, V., R. R. Sealock, R. H. Sifferd, O. Kamm & I. W. Grote. 1933. Proc. Am. Soc. Biol. Chem., J. Biol. Chem. **100:** xciv.
22. Irving, G. W., Jr., H. M. Dyer & V. du Vigneaud. 1941. J. Am. Chem. Soc. **63:** 503.
23. Sealock, R. R. & V. du Vigneaud. 1935. J. Pharmacol. Exptl. Therap. **54:** 433.
24. Kamm, O. 1928. Science. **67:** 199.
25. du Vigneaud, V., G. W. Irving, Jr., H. M. Dyer & R. R. Sealock. 1938. J. Biol. Chem. **123:** 45.
26. Irving, G. W., Jr. & V. du Vigneaud. 1938. J. Biol. Chem. **123:** 485.
27. Cohn, M., G. W. Irving, Jr. & V. du Vigneaud. 1941. J. Biol. Chem. **137:** 635.
28. Craig, L. C. 1944. J. Biol. Chem. **155:** 519.
29. Livermore, A. H. & V. du Vigneaud. 1949. J. Biol. Chem. **180:** 365.
30. Turner, R. A., J. G. Pierce & V. du Vigneaud. 1951. J. Biol. Chem. **191:** 21.
31. Pierce, J. G. & V. du Vigneaud. 1950. J. Biol. Chem. **182:** 359; **186:** 77.
32. Popenoe, E. A., H. C. Lawler & V. du Vigneaud. 1952. J. Am. Chem. Soc. **74:** 3713.
33. Pierce, J. G., S. Gordon & V. du Vigneaud. 1952. J. Biol. Chem. **199:** 929.
34. Light, A. & V. du Vigneaud. 1958. Proc. Soc. Exptl. Biol. Med. **98:** 692.
35. Mueller, J. M., J. G. Pierce, H. Davoll & V. du Vigneaud. 1951. J. Biol. Chem. **191:** 309.
36. Turner, R. A., J. G. Pierce & V. du Vigneaud. 1951. J. Biol. Chem. **193:** 359.
37. Davoll, H., R. A. Turner, J. G. Pierce & V. du Vigneaud. 1951. J. Biol. Chem. **193:** 363.
38. Mueller, J. M., J. G. Pierce & V. du Vigneaud. 1953. J. Biol. Chem. **204:** 857.
39. Ressler, C., S. Trippett & V. du Vigneaud. 1953. J. Biol. Chem. **204:** 861.
40. du Vigneaud, V., C. Ressler & S. Trippett. 1953. J. Biol. Chem. **205:** 949.
41. Popenoe, E. A. & V. du Vigneaud. 1953. J. Biol. Chem. **205:** 133.
42. Popenoe, E. A. & V. du Vigneaud. 1954. J. Biol. Chem. **206:** 353.
43. Lawler, H. C. & V. du Vigneaud. 1953. Proc. Soc. Exptl. Biol. Med. **84:** 114.
44. du Vigneaud, V., H. C. Lawler & E. A. Popenoe. 1953. J. Am. Chem. Soc. **75:** 4880.
45. Tuppy, H. 1953. Biochim. et Biophys. Acta. **11:** 449.
46. Acher, R. & J. Chauvet. 1954. Biochim. et Biophys. Acta. **14:** 421.
47. Sifferd, R. H. & V. du Vigneaud. 1935. J. Biol. Chem. **108:** 753.
48. du Vigneaud, V. & G. L. Miller. 1936. J. Biol. Chem. **116:** 469.
49. Gordon, S. & V. du Vigneaud. 1953. Proc. Soc. Exptl. Biol. Med. **84:** 723.
50. du Vigneaud, V., C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis & S. Gordon. 1953. J. Am. Chem. Soc. **75:** 4879.
51. Ressler, C. & V. du Vigneaud. 1954. J. Am. Chem. Soc. **76:** 3107.
52. Swan, J. M. & V. du Vigneaud. 1954. J. Am. Chem. Soc. **76:** 3110.
53. Katsoyannis, P. G. & V. du Vigneaud. 1954. J. Am. Chem. Soc. **76:** 3113.
54. du Vigneaud, V., C. Ressler, J. M. Swan, C. W. Roberts & P. G. Katsoyannis. 1954. J. Am. Chem. Soc. **76:** 3115.

55. Bodanszky, M. & V. du Vigneaud. 1959. J. Am. Chem. Soc. **81:** 2504.
56. Bartlett, M. F., A. Jöhl, R. Roeske, R. J. Stedman, F. H. C. Stewart, D. N. Ward & V. du Vigneaud. 1956. J. Am. Chem. Soc. **78:** 2905.
57. du Vigneaud, V., M. F. Bartlett & A. Jöhl. 1957. J. Am. Chem. Soc. **79:** 5572.
58. Bodanszky, M. & V. du Vigneaud. 1959. Nature. **183:** 1324.
59. Bodanszky, M. & V. du Vigneaud. 1959. J. Am. Chem. Soc. **81:** 5688.
60. Bodanszky, M. 1955. Nature. **175:** 685.
61. Boissonnas, R., St. Guttmann, P.-A. Jaquenoud & J.-P. Waller. 1955. Helv. Chim. Acta. **38:** 1491.
62. Rudinger, J., J. Honzl & M. Zaoral. 1956. Chem. Listy. **50:** 288. Collection Czechoslov. Chem. Communications. **21:** 202.
63. Bodanszky, M., M. Szelke, E. Tomorkeny & E. Weiss. Unpublished work; *cf.* L. Gyermek and Gy. Fekete. Experientia. **11:** 238.
64. Velluz, L., G. Amiard, V. Bartos, B. Goffinet & R. Heymes. 1956. Bull. soc. chim. France. **:** 1464.
65. Beyerman, H. C., J. S. Bontekoe & A. C. Koch. 1959. Rec. trav. chim. **78:** 935.
66. du Vigneaud, V., D. T. Gish, P. G. Katsoyannis & G. P. Hess. 1958. J. Am. Chem. Soc. **80:** 3355.

Part I. Isolation and Characterization

ION-EXCHANGE CHROMATOGRAPHY OF INSULIN AND OTHER PROTEINS IN BUFFERS CONTAINING UREA*

R. David Cole and Leticia Mendiola
Department of Biochemistry, University of California, Berkeley, Calif.

The purification of proteins may be divided into stages. In the first stage closely related families of proteins may be isolated from each other by various means, including chromatography, using stepwise changes of *p*H or ionic strength. A second stage might be the further fractionation of one of these families by use of chromatography involving much smaller steps in the change of *p*H or ionic strength or even involving a continuous *p*H or ionic strength gradient. In the last stage a nearly pure protein is scrutinized for signs of inhomogeneity.

As this final stage in the purification of a protein is approached, the most advantageous form of chromatography is probably elution analysis in which a buffer of uniform composition is used throughout the entire chromatogram.[1] One such chromatographic system uses the carboxylic ion-exchange resin Amberlite IRC-50. However, elution analysis using this resin has been successfully accomplished only when applied to neutral or basic proteins.[1] The problem with other proteins is that, in buffers having a *p*H low enough to bring about an appreciable binding of the nonbasic protein to the resin, the protein is more or less irreversibly bound and cannot be released from the resin until the *p*H or the ionic strength of the buffer has been significantly increased. Schmid *et al.*[2] have exploited this all-or-none type of binding in extending the use of Amberlite IRC-50 chromatography to nonbasic proteins. These workers allow a mixture of proteins to bind to the resin at low *p*H values and then wash the proteins from the resin successively by stepwise increases of *p*H. While such an approach has a valuable place in protein purification, particularly in the earlier stages of purification when the isolation of families of closely related proteins may be desirable, elution by sudden changes of *p*H or ionic strength fails to subject proteins to the resolving power characteristic of elution analysis.

The possibility of chromatographing nonbasic proteins on Amberlite IRC-50 by elution analysis still exists, and hope of its accomplishment may be based on the work of Boardman and Partridge.[3] These authors suggest that the relatively irreversible binding of the protein to the resin in solutions below *p*H 6 is in large part due to hydrogen bonding. According to their explanation, when the *p*H of the buffer is lowered enough to make the charge on the protein suitable for ion exchange, the buffer causes appreciable protonation of the carboxyl groups of the resin which then bind each protein molecule through hydrogen bonds. Such bonds are likely to form multipoint attachments between the protein and the resin with the result that the protein cannot move down the column unless all the protein-resin bonds are broken simultaneously.

* The work described in this paper was supported in part by Grant A2691 from the National Institute of Arthritis and Metabolic Diseases, Public Health Service, Bethesda, Md.

A buffer with *p*H high enough to avoid hydrogen bonding would render the protein unsuitable for ion exchange, while a buffer with *p*H low enough for the ion-exchange equilibrium would result in the hydrogen-bonding problem. Clearly it is difficult or impossible to find conditions that would avoid hydrogen bonding and simultaneously allow ion exchange of a freely reversible kind.

At least two possibilities exist for overcoming the problem of multiple hydrogen bonding. One possibility is to reduce the concentration of carboxyl groups on the resin. Such an approach can be seen in the work of Boardman, who prepared a new resin on which he recently chromatographed insulin.[4] Resins containing a lower concentration of carboxyl groups could also be prepared as carboxymethyl cellulose.[5] A second possibility for overcoming hydrogen bonding is to use the Amberlite IRC-50 resin but to attempt the disruption of the multiple hydrogen bonding by the use of chemicals such as urea. If the hydrogen bonding could be eliminated in this way, simple ion exchange could take place and chromatography could be accomplished using a single buffer throughout any one experiment. The work to be reported represents a study of this latter approach and its application to the chromatography of model nonbasic proteins, insulin, glucagon, and prolactin.

Experiments

Protein samples. Bovine crystalline insulin,* porcine crystalline insulin,† amorphous porcine samples,‡ and crude preparations of porcine insulin§ were used. Other samples of crude insulin, prepared essentially as described by Best *et al.*[6] were also used, along with samples of glucagon.¶ The prolactin used was prepared as described by Cole and Li.[7]

Assays. The assay of insulin was carried out on fasted rabbits, observing the blood glucose levels by use of the Somogyi method.[8] The assay of prolactin was for the crop-sac-stimulating potency as described by Riddle *et al.*[9]

Preparation of buffers. Solutions of 8 *M* urea or other stated concentrations were prepared and deionized by passing them through columns of mixed-bed ion-exchange resin (Amberlite MB-1). This deionized urea was then used to dissolve the salts in preparing a solution of 0.13 *M* NaH_2PO_4-8 *M* urea and in making a solution of 0.13 *M* Na_2HPO_4-8 *M* urea. Small amounts of the 0.13 *M* Na_2HPO_4-8 *M* urea solution were added to the 0.13 *M* NaH_2PO_4-8 *M* urea until the desired *p*H (as read using a glass electrode) was reached. To 1 l. of the NaH_2PO_4 solution, Na_2HPO_4 solution was added to the extent of 60 ml. for *p*H 6.00, 70 ml. for *p*H 6.05, or 80 ml. for *p*H 6.10. It may be observed that the concentration of Na^+ varies slightly from a buffer of one *p*H to a buffer of another *p*H.

Preparation of resin and operation of columns. The resin used was either Amberlite IRC-50 (XE-64) screened through 200-mesh gauze, or Amberlite IRC-50 (XE-97). The columns were prepared and used as described by Hirs

* Eli Lilly and Co., Indianapolis, Ind., Lots 535664 and 693502, and International Standard supplied by British Drug Houses, Ltd., London, England, Batch 2189.
† Eli Lilly and Co., Lot 723603.
‡ Eli Lilly and Co., Lot 200-1B-17.
§ Eli Lilly and Co.
¶ Eli Lilly and Co., Lots 258-234B-54-2 and 258-234B-167-1.

et al.[10] except that the buffers contained urea. The chromatography was carried out in cold rooms at 3 to 5° C. at a flow rate of 3 to 5 ml. per hour. Insulin was applied in amounts of 1 mg. to 100 mg. to columns of 0.90 × 20 cm. or 0.64 × 30 cm.

Analysis of effluent fractions and isolation of protein. Frequently the whole fraction (0.5 to 1.0 ml.) was used for analysis in the test tube in which it was collected; at other times aliquots were analyzed. The assay for protein was carried out using the procedure described by Lowry *et al.*[11]

Insulin can be recovered from the effluent fraction by precipitation with acetone, followed by washing with water and drying with alcohol and ether. Alternatively, a faster forming precipitate may be obtained simply by diluting the effluent with 10 volumes of water; this precipitate may be collected, washed, and dried with ethanol and ether in 70 to 75 per cent yield. In this latter procedure, another 15 to 17 per cent yield may be recovered from the supernatant of the diluted effluent fraction by saturating this supernatant with Na_2SO_4.

Glucagon and prolactin may be isolated by the procedures used for insulin, or prolactin may be dialysed and lyophilized.

Results

If urea was to be used to disrupt hydrogen bonding while allowing ion exchange to take place, it was clear that urea must not interfere with the ion-exchange process itself. To test this condition, the chromatograph of pancreatic ribonuclease, a basic protein, was carried out under the conditions described by Hirs *et al.*[10] except that the phosphate buffers were made 8 *M* with respect to urea. The chromatographic pattern (FIGURE 1) obtained was essentially the same as the pattern observed in the absence of urea.

Having established that urea had no disturbing effects on ion exchange in the case of a basic protein, the study was extended to a model, nonbasic protein, insulin. The chromatography of insulin was attempted at various *p*H values using buffers that were 0.13 *M* phosphate and 8 *M* urea. The relative elution volumes of the major component are plotted as a function of *p*H in FIGURE 2. The fact that this curve does not exhibit a sharp break from a region of no binding to a region of total binding shows that the binding of protein by the resin occurs in a freely reversible manner over a reasonable range of *p*H, near *p*H 6. That ion exchange played a major role in the chromatography under these conditions (that is, phosphate-urea buffers of *p*H 6) may be inferred from the sensitivity of the chromatography to ionic strength. In FIGURE 3 it may be observed that in 0.17 *M* phosphate the elution volume of the main peak of insulin was much less than the elution volume observed when 0.13 *M* phosphate was used.

Optimal conditions for the chromatography of insulin appear to approximate 0.13 *M* phosphate-7 *M* urea at *p*H 6.0. Under these conditions resolution was entirely satisfactory, and the insulin was quantitatively eluted (97 to 105 per cent based on the color obtained). It is of interest to identify the 3 peaks (peak 1, the peak at the hold-up volume; peak 2, the intermediate peak; and peak 3, the slowly moving main component) of the chromatogram of crystalline

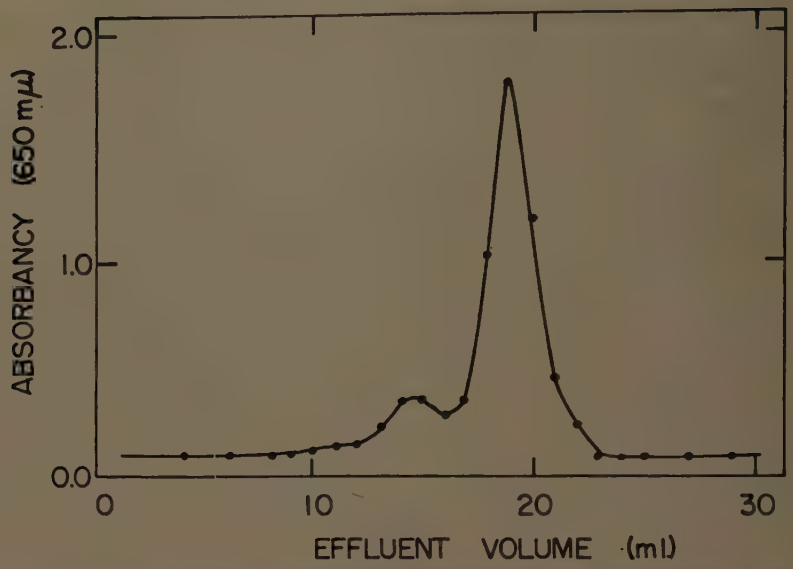


FIGURE 1. A chromatogram of ribonuclease obtained using 0.2 *M* phosphate-8 *M* urea buffer at *p*H 6.47 on a column of Amberlite IRC-50, 0.9 × 20 cm.

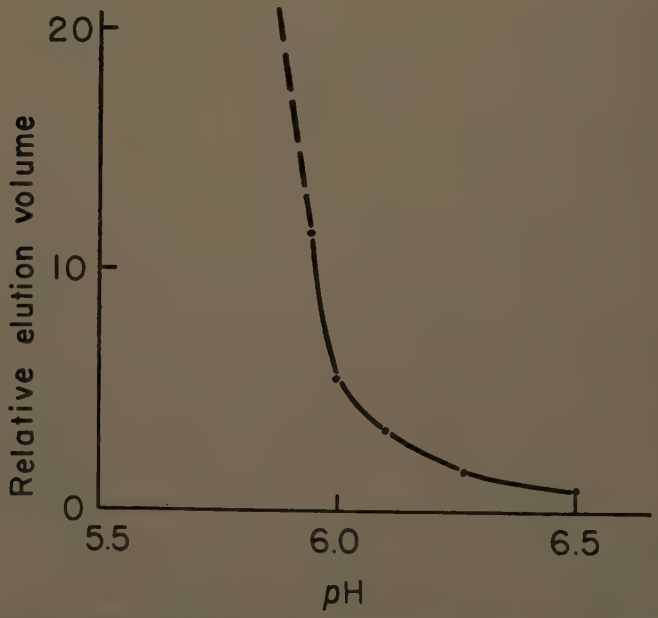


FIGURE 2. The relative elution volume (in arbitrary units) of the main component of insulin in 0.13 *M* phosphate-8 *M* urea at various *p*H values. These volumes were measured by chromatography on columns of Amberlite IRC-50, 0.9 × 20 cm. or 0.64 × 30 cm. FIGURES 2 through 8, 10, 11, 20, and 21 in this paper are reproduced by permission from the *Journal of Biological Chemistry*.

bovine insulin, especially as Harfenist and Craig[12] had demonstrated by counter-current distribution the presence of 2 components in significant amounts. Harfenist[13] found that the smaller of their two peaks, which they called "peak B," differed from the main component only in having one less amide residue per molecule (their minor component has been called desamidoinsulin). One of the samples that they studied was an amorphous sample of porcine insulin that contained a large proportion of desamidoinsulin; when this same sample was

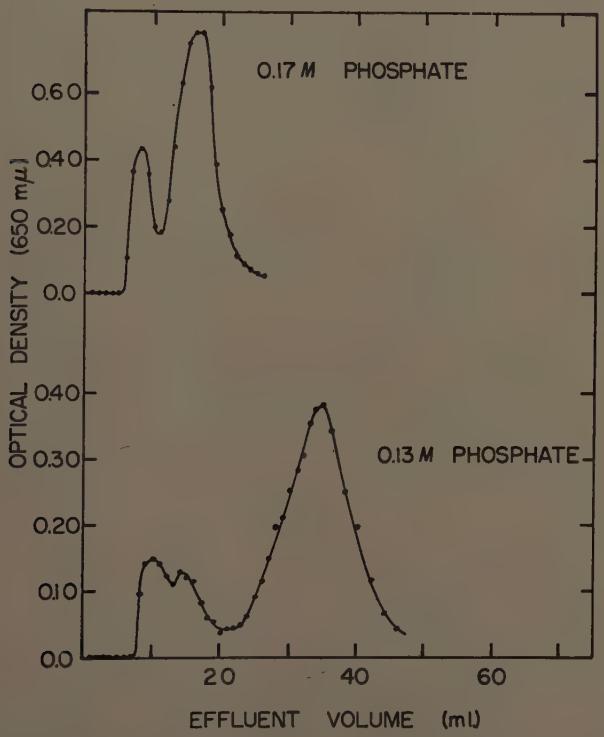


FIGURE 3. The chromatography of crystalline bovine insulin on Amberlite IRC-50 at *p*H 6.0 in 6 *M* urea buffers of different ionic strengths. The columns were 0.64 × 28 cm.

submitted to chromatography in phosphate-urea buffer, the chromatographic component seen in especially large proportion was peak 1 (FIGURE 4). This indicated that the fastest moving peak on the chromatogram corresponded at least in part to the desamidoinsulin of Harfenist and Craig, and this indication is borne out by FIGURE 5, which presents a chromatogram of desamido-insulin purified by counter-current distribution.

Having located desamidoinsulin in the first peak of the chromatogram, we submitted the main counter-current fraction to chromatography with the results given in FIGURE 6. The pattern shown reveals that, while the main counter-current fraction corresponds largely to the main chromatographic component (peak 3), it also contains a significant amount of peak 2. The nature

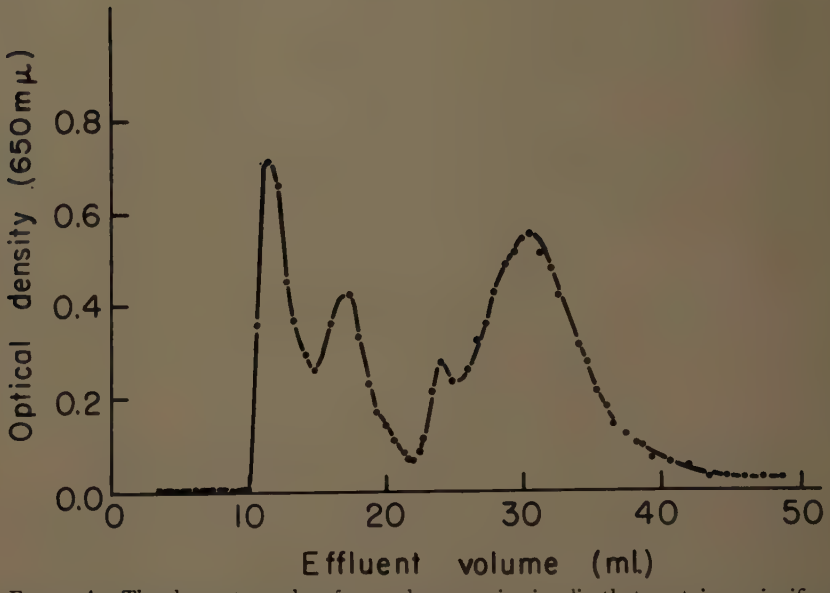


FIGURE 4. The chromatography of amorphous porcine insulin that contains a significant amount of desamidoinsulin. The column was 0.9 × 20 cm. of Amberlite IRC-50 and the eluting buffer was 0.13 *M* phosphate-7 *M* urea at *p*H 6.0.

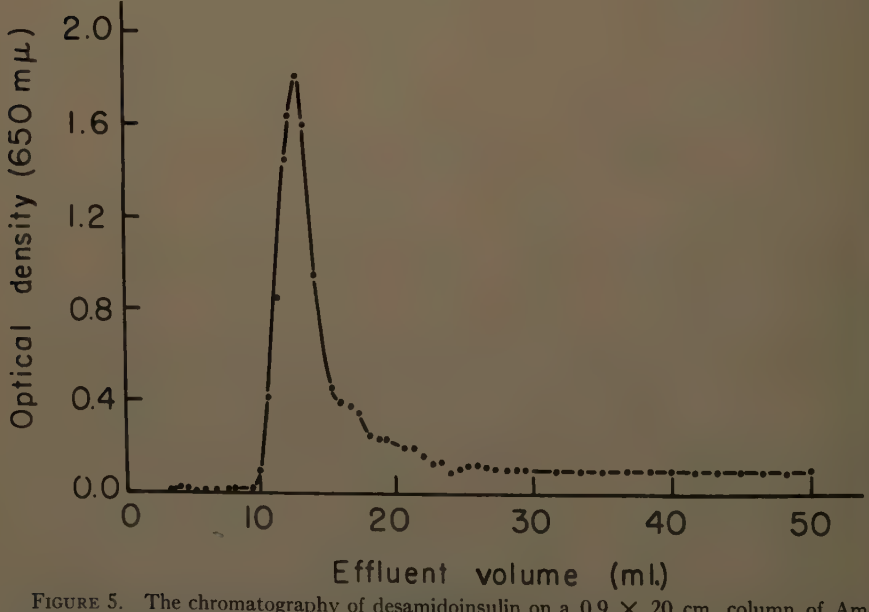


FIGURE 5. The chromatography of desamidoinsulin on a 0.9 × 20 cm. column of Amberlite IRC-50, eluting with 0.13 *M* phosphate-7 *M* urea at *p*H 6.0.

of peak 2 is not clear. That the presence of peak 2 does not arise as the result of a reversible equilibrium with the main component was demonstrated by the rechromatography of peak 3 (FIGURE 7). That neither peak 1 nor peak 2 arose, one from the other, was shown by their separate rechromatography. The results of these experiments are given in FIGURE 8. It must be concluded, then, that peak 2 corresponds to a discrete stable component of these preparations of insulin.

While material corresponding to peak 2 was originally present in crystalline insulin and in fractions isolated from counter-current distribution, it was also possible to produce an increase in peak 2 by incubation of the insulin in phosphate-urea buffer at room temperature before chromatography. This phe-

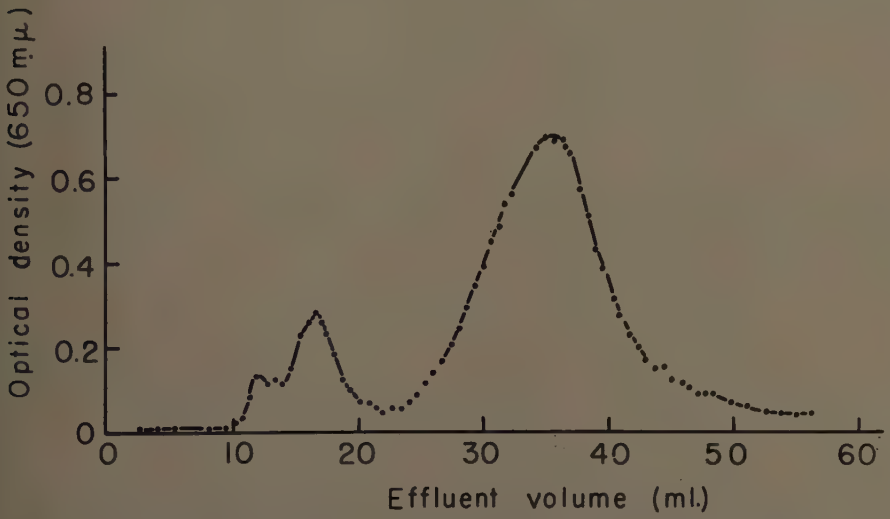


FIGURE 6. The chromatography of the main counter-current fraction of insulin on a 0.9 × 20 cm. column of Amberlite IRC-50, eluting with 0.13 *M* phosphate-7 *M* urea at *p*H 6.0.

nomenon was discovered when chromatography was initially performed at room temperature. Although the initial chromatography at room temperature gave a chromatographic pattern that was indistinguishable from the pattern obtained at 3 to 5° C., rechromatography of material from the chief band of these two types of chromatograms showed a difference. If the initial experiment had been carried out at room temperature, the rechromatography would have shown that a part of the insulin had been transformed to material giving a peak coinciding with peak 2. This may be seen in FIGURE 9, which presents the rechromatography of eluate taken from the main peak of a previous chromatogram run at room temperature (the protein had been in contact with 0.13 *M* phosphate-8 *M* urea, *p*H 6.0 buffer, for about 2 days at room temperature when this chromatogram was started). In contrast, if the initial experiment had been done at 4° C., rechromatography of peak 3 material would have shown a single peak at the appropriate elution volume as was shown in FIGURE 7 (for this reason, and because the buffers are more stable in the cold, all of the

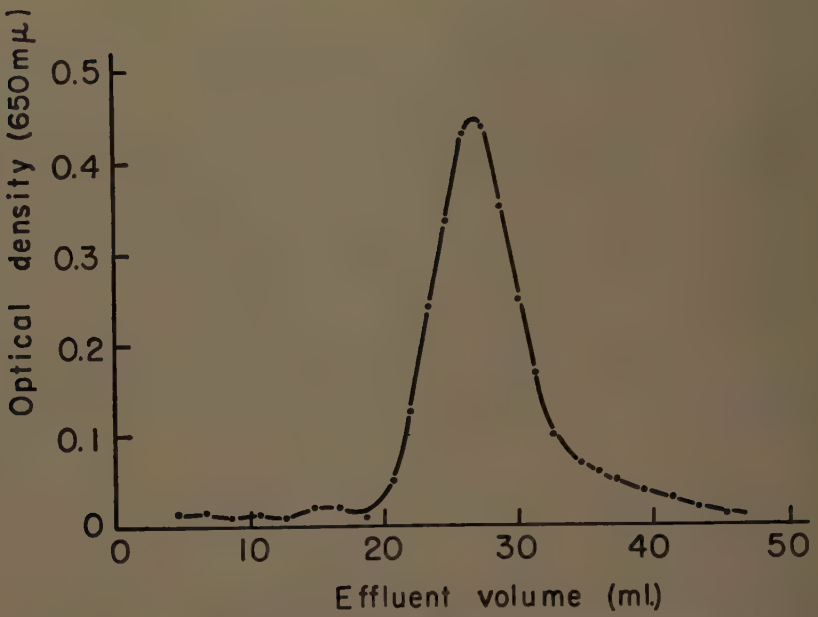


FIGURE 7. The rechromatography of material taken from the major peak of a previous chromatogram of bovine insulin. The column was 0.9 × 20 cm. of Amberlite IRC-50, and the buffer was 0.13 *M* phosphate-7 *M* urea, *p*H 6.0.

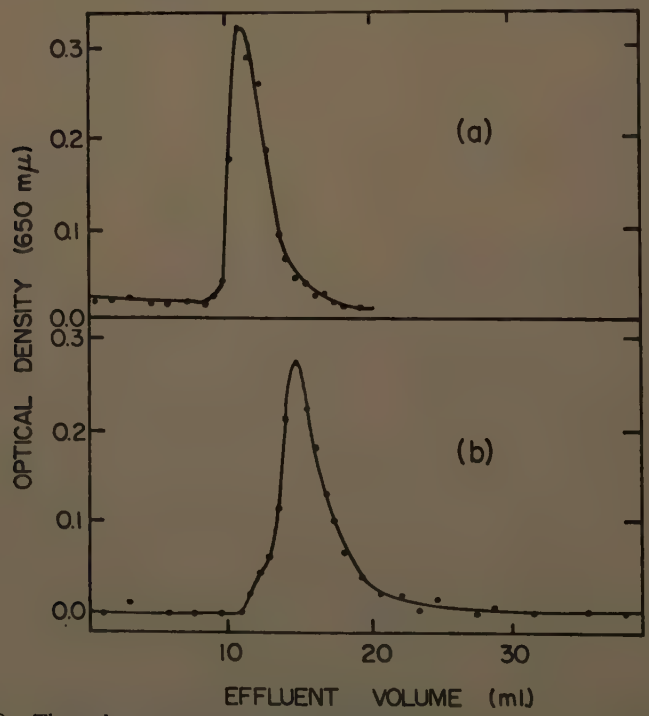


FIGURE 8. The rechromatography of material taken from the minor peaks of a previous chromatogram of bovine insulin. The column was 0.9 × 20 cm. of Amberlite IRC-50 and the buffer was 0.13 *M* phosphate-7 *M* urea, *p*H 6.0. (*a*) Material taken from the fastest running peak (peak 1) of a previous chromatogram. (*b*) Material taken from the intermediate peak (peak 2) of a previous chromatogram.

other chromatography reported here was carried out at 3 to 5° C.). The production of material that chromatographs as peak 2, by prior incubation of peak 3 material in phosphate-urea buffer, suggests that the presence of peak 2 material in crystalline insulin and in the counter-current fractions may be due to subtle transformations of the main component during isolation and handling. The transformation of insulin in warm, concentrated urea has been studied by Bischoff and Bakhtiar,[14] and it may be that peak 2 in these chromatograms represents the initial intermediate in the series of reactions studied by these workers.

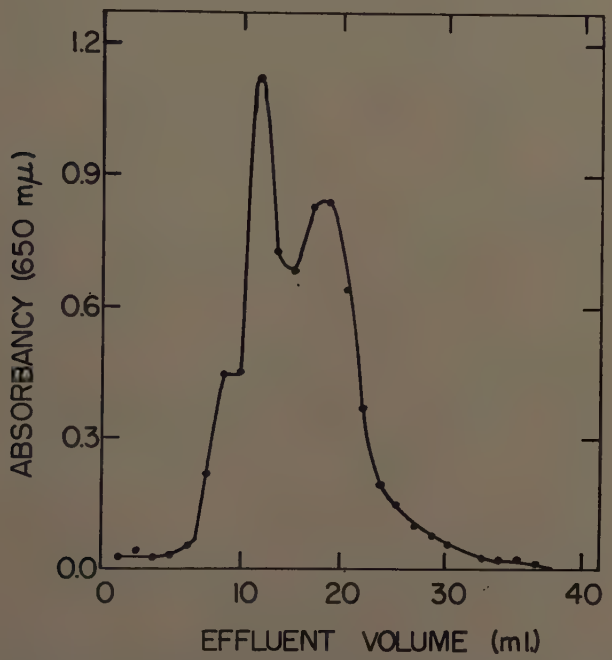


FIGURE 9. The rechromatography of eluate from the main peak of a chromatogram run at room temperature in 0.13 *M* phosphate-8 *M* urea, *p*H 6.0. The eluate was applied to a 0.9 × 20 cm. column of Amberlite IRC-50 after being in contact with the buffer for 2 days.

It is also pertinent to point out that Carpenter and Hess[15] and Chrambach,[16] using partition chromatography, have found a component of insulin that is an active transformation product derived from the main fraction of insulin by hydrochloric acid treatment. This material was submitted to ion-exchange chromatography, and the chromatogram showed a peak that coincided with peak 2. The similarity of the acid-transformed material to the product of phosphate-urea transformation is interesting and surprising, and the interest is extended by the finding that material isolated from peak 2 of the ion-exchange chromatograms also was active. Chrambach and Carpenter are investigating the nature of their transformation product.

To make certain that peak 2 was not peculiar to a few preparations of insulin, chromatograms were obtained using crystalline bovine insulin from different

sources, porcine as well as bovine insulin was investigated, and chromatograms were prepared using both crystalline and amorphous hormone. In every case peak 2 was present in a significant amount. The only instances in which chromatograms failed to show peak 2 was the chromatography of material isolated from peak 3 as described above or isolated from the main chromatographic component of Chrambach.[16]

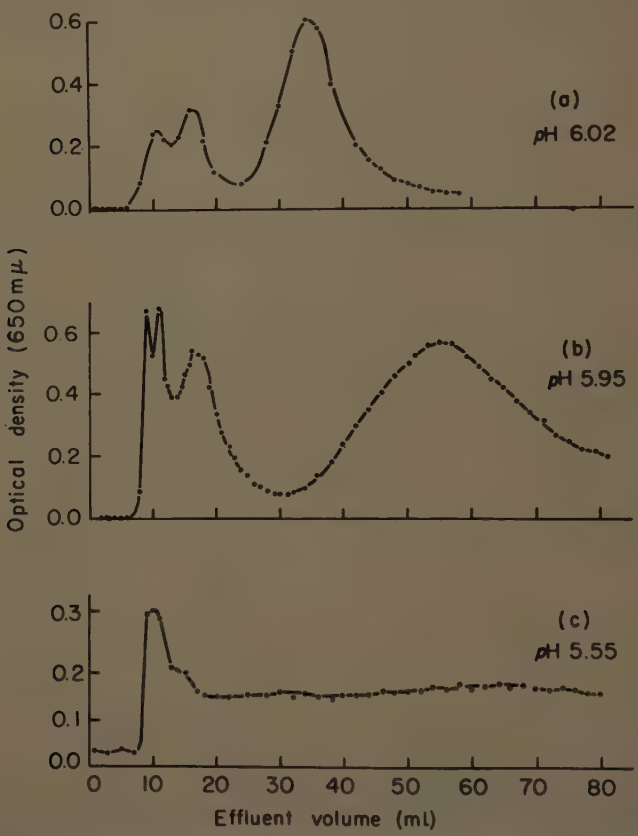


FIGURE 10. The chromatography of crystalline bovine insulin on Amberlite IRC-50, eluting with 0.13 *M* phosphate-7 *M* urea buffers of various *p*H values. The columns were 0.9 × 20 cm. or 0.64 × 30 cm.

In the interests of applying this chromatographic system to proteins in general, it was gratifying to find that ion-exchange chromatography could be accomplished under the conditions originally postulated. The protein recovery, the reproducibility, the capacity, and the resolving power were entirely satisfactory. The resolving power was in fact far superior to expectation. The fact that a third active form of insulin was not revealed in the chromatographic system of Boardman[4] or in that of Samsonov and Faddeeva[17] may have been the result of their having used different types of insulins as starting materials, but more likely the difference between their results and those presented

here was due to the different resins used. In the present case the unexpectedly high power of resolution probably depends on the combined action of ion exchange and hydrogen bonding. There might thus be added to the usual power of ion exchange an increased power of discrimination among proteins according to their shapes, at least to the extent that the molecular shapes determine the quality and quantity of hydrogen-bonding sites available to the resin.

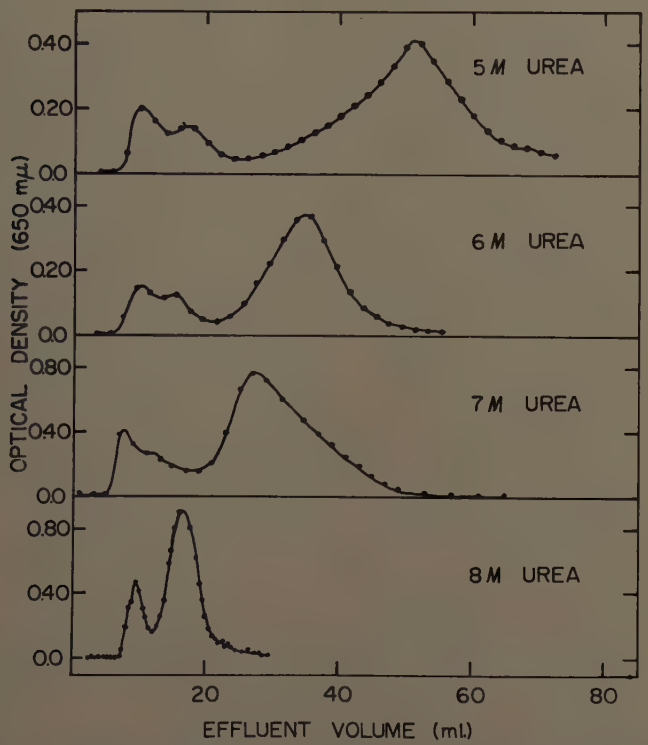


FIGURE 11. The chromatography of crystalline bovine insulin on Amberlite IRC-50 eluting with *p*H 6.0 buffers which were 0.13 *M* phosphate and different urea concentrations. These columns were 0.64 × 30 cm. The shape of the curve obtained using 7 *M* urea appears skewed because the column was slightly overloaded (100 mg. of crystalline insulin was applied). The other columns received 5 mg. insulin.

The involvement of hydrogen bonding along with ion exchange may be suggested on the basis of the series of chromatograms shown in FIGURE 10. As the *p*H of the phosphate-urea buffers was lowered, maintaining all other conditions, the various peaks of the chromatographic pattern were not only retained longer, as would be expected on the basis of ion exchange, but the peaks became increasingly and severely spread, suggesting a severe surface binding like hydrogen bonding. Further support for the concept that hydrogen bonds are still playing a significant part comes from the data presented in FIGURE 11. In this figure there is presented a series of chromatograms obtained with the

use of 0.13 *M* phosphate buffers of *p*H 6.0 in which the concentration of urea was varied. It may be observed that the lowering of the urea concentration increased the retention and spreading of the protein bands.

While it is easy to postulate the involvement of both ion exchange and hydrogen bonding, it is not easy to draw a detailed picture of the mechanisms of hydrogen bonding. Of the various ways in which urea might be acting in the chromatography of proteins three general effects may be considered. First, the urea may disrupt hydrogen bonds between resin and protein, in which case increased urea concentrations would decrease retention. Second, disruption of intrachain hydrogen bonds of the protein would be observed as an increase of retention as the urea concentration was increased because unfolding of the molecule would expose more of the protein's hydrogen bonding sites to the surface of the resin. Finally, urea may be acting by dissociating aggregates of the protein molecule: for example, insulin forms aggregates of molecular weight 36,000 or more in neutral aqueous solutions, but would be expected to dissociate as urea is added, perhaps decreasing to a molecular weight of 6000.[18] These smaller units would have less tendency to form multipoint attachments to the resin; therefore we predict that increasing urea concentration would cause decreased retention. In the chromatograms already shown, the retention volume of insulin was decreased when the urea concentration was increased; accordingly the predominant effects of urea would seem to be the dissociation of protein aggregates, the disruption of protein-resin hydrogen bonds, or both.

In adapting this kind of chromatography to other proteins, however, any or all of the three general effects might be expected. While the establishment of conditions for ion-exchange chromatography of proteins is usually at least partly empirical, in the work reported here it is even more so. This is because the extra factor of urea concentration is involved. While increased empiricism is not usually desirable, the extra factor does result in a highly desirable increased sensitivity. While it may be predicted in a general way that all proteins will show increased retention with decreasing *p*H or ionic strength, it may not be predicted safely that all the components of a protein mixture will respond to the variation of urea concentration in the same way. For example, one component of a protein mixture might be eluted from a column more and more rapidly as the urea concentration is increased up to the point at which the protein molecule unfolds, and then be eluted progressively (or suddenly) more slowly. Another component of that same protein mixture might not undergo any change of molecular folding over the range of urea concentration employed and so would show a progressive decrease in elution volume as the urea concentration increases. Clearly the relative positions of the two protein peaks and even the order of elution of the two proteins might change with changing urea concentration. For this reason the adaptation of this type of chromatography to previously untried proteins may require a rather extensive study of the effects of urea in that particular situation. Some of these problems in predicting the chromatographic behavior of different proteins were illustrated when an attempt was made to isolate insulin from different animal species.

A study of the isolation of insulin from laboratory animals was undertaken because the prospect of obtaining a method for the preparation and analysis

of insulin in various animals, especially laboratory animals, was attractive as a tool for the study of protein synthesis and diabetogenesis as well as for a study of the comparative biochemistry of insulin. First we tested the applicability of the chromatographic system to crude preparations of insulin by studying samples of crude pork insulin. FIGURE 12 shows a chromatogram of crystalline pork insulin in 8 *M* urea that was run first. The pattern is very similar to that of beef insulin. Then a sample of insulin partially purified (16 U./mg.) was investigated with the results given in FIGURE 13. It may be seen that the insulin peak was adequately isolated by the chromatographic system and that its chromatographic behavior was undistorted. FIGURE 14 shows a chromatogram of a preparation that was only about 10 per cent insulin (2 U./mg.), and still the chromatography in 8 *M* urea was able to accomplish the purification

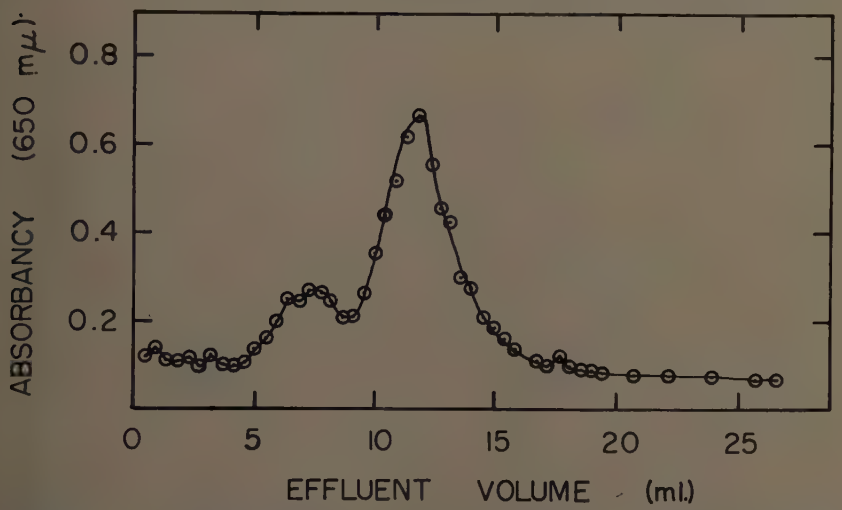


FIGURE 12. The chromatography of crystalline pork insulin on Amberlite IRC-50, in *p*H 6.0 buffers, 0.13 *M* phosphate-8 *M* urea. The column was 0.64 × 30 cm.

of the insulin, which was eluted at about 13 to 14 ml. When 7 *M* urea was used, the insulin peak was retained longer and resolution was improved. Being somewhat uncertain of the commercial process used in obtaining these crude extracts, an extract of beef insulin was prepared that was about 4 per cent pure (1 U./mg.) and observed in a chromatogra n in 7 *M* urea, as shown in FIGURE 15. The size of this peak corresponded to the amount of insulin expected and was well resolved from the other components of the system. Therefore, since the purification procedure appeared to be satisfactory, it was applied to the isolation of insulin from rabbit pancreas with the results shown in FIGURE 16. FIGURE 16*a* shows a chromatogram of crude rabbit insulin; assays of the insulin activity identified the shoulder on the peak of unretained protein as rabbit insulin. In order to be sure that some unknown factor in the crude preparation was not pulling the insulin out of position, some crystalline beef insulin was mixed with the crude rabbit insulin, and the chromatogram shown in FIGURE 16*b* was obtained by using that mixture. Clearly if any factor distorts

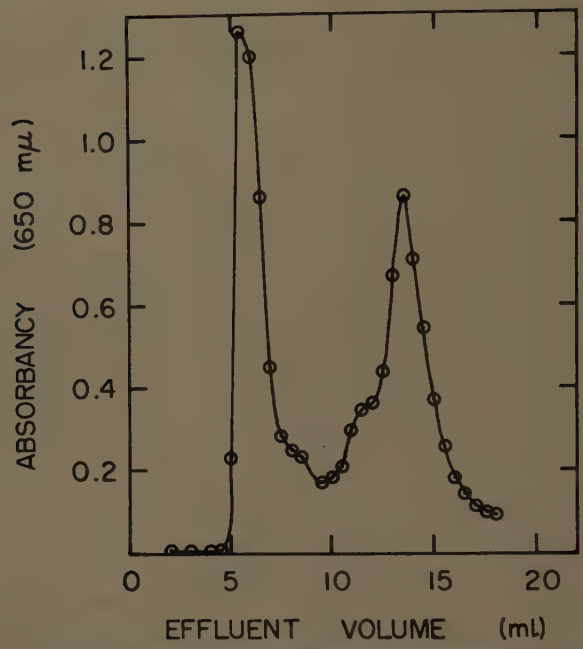


FIGURE 13. The chromatography of partially purified pork insulin (16 U./mg.) in 0.13 *M* phosphate-8 *M* urea, at *p*H 6.0. A column of Amberlite IRC-50, 0.64 × 30 cm. was used.

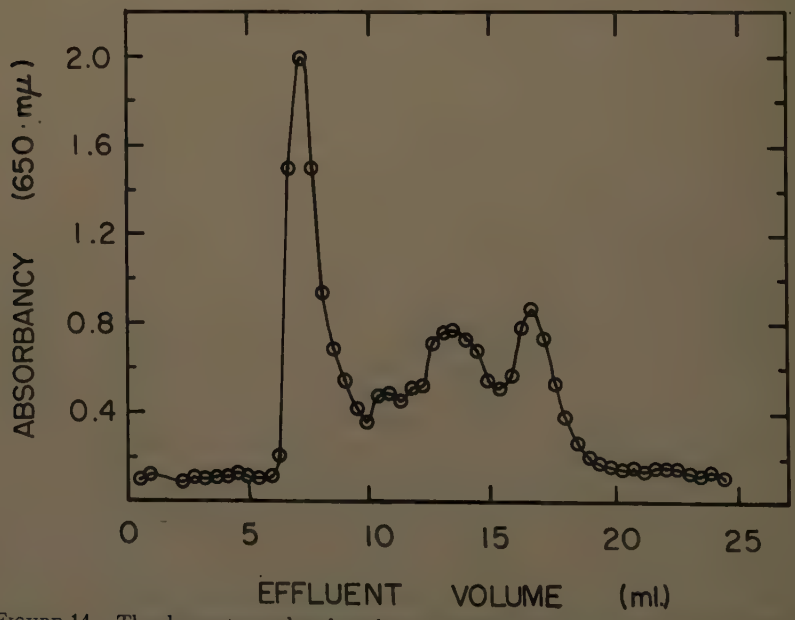


FIGURE 14. The chromatography of crude pork insulin (2 U./mg.) in 0.13 *M* phosphate-8 *M* urea at *p*H 6.0 on a 0.64 × 30 cm. column of Amberlite IRC-50.

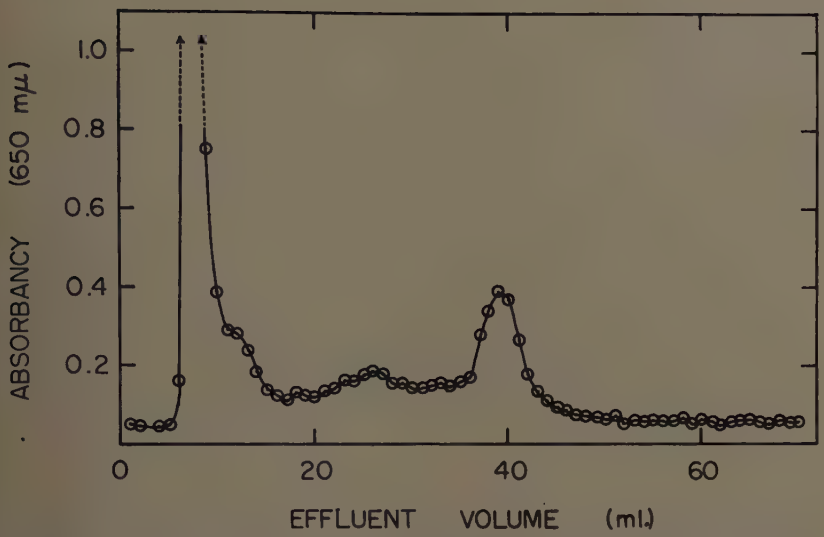


FIGURE 15. A chromatogram of crude beef insulin (1 U./mg.) in 0.13 *M* phosphate-7 *M* urea at *p*H 6.0 on a 0.64 × 30 cm. column of Amberlite IRC-50.

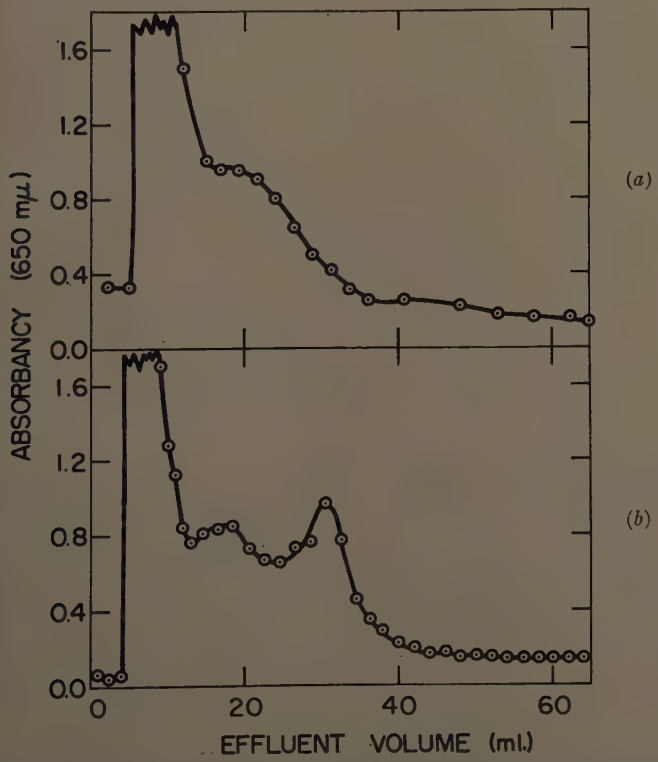


FIGURE 16. (*a*) A chromatogram of crude rabbit insulin (1 U./mg.) in 0.13 *M* phosphate-7 *M* urea at *p*H 6.0 on a 0.64 × 30 cm. column of Amberlite IRC-50. (*b*) A chromatogram of crude rabbit insulin mixed with crystalline beef insulin run under the same conditions as above.

the chromatogram of rabbit insulin, it does not affect beef insulin. Thus the difference in the elution volumes of rabbit and ox insulins in 7 *M* urea seems real and is very interesting. Perhaps even more intriguing was the finding that the response of the two insulins to urea was different. This was recognized when rabbit insulin was submitted to chromatography in 5 *M* urea. The chromatogram obtained is shown in FIGURE 17. Here the insulin was eluted at about 50 ml., just as bovine insulin was at this concentration of urea. This, then, is a case in which the relative elution volumes varied at different urea concentrations.

In contrast to this case is the comparison of pork and beef insulins. In 10 pairs of chromatograms in which beef insulin was compared to pork insulin

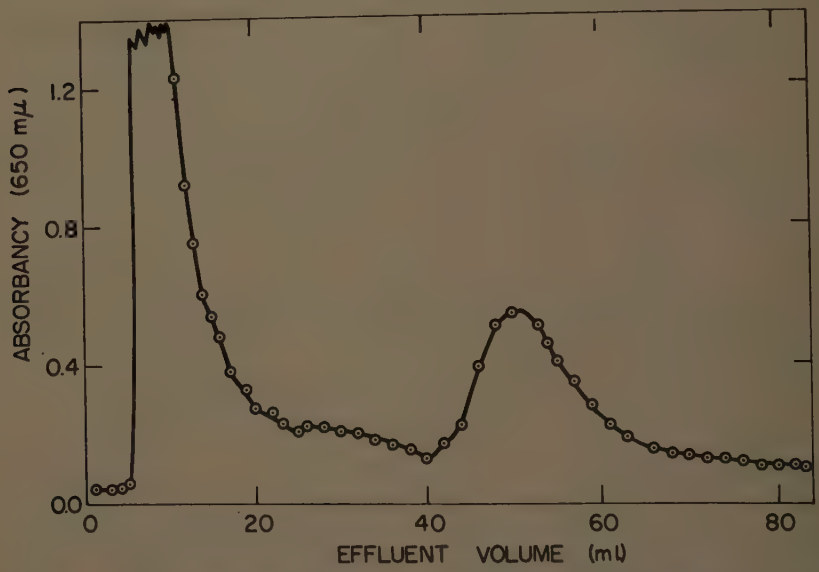


FIGURE 17. The chromatography of crude rabbit insulin in 0.13 *M* phosphate-5 *M* urea at *p*H 6.0 on a 0.64 × 30 cm. column of Amberlite IRC-50.

throughout the range 5 *M* to 8 *M* urea, the beef hormone was always found to be retained about 20 per cent longer than the pork hormone. No conditions were found that gave better resolution than observed in FIGURE 18, in which a mixture of pork and beef insulins is compared to beef insulin alone. The mixture gives the pattern that would be calculated if appropriate independent chromatograms of the two hormones were added together.

Still another type of chromatographic behavior was observed by submitting crude preparations of chicken insulin to chromatography in 7 *M* urea with the results shown in FIGURE 19. The insulin activity lies in the rear portion of the main peak, and when a lower concentration of urea (5 *M*) was used for the chromatography of crude chicken insulin, the pattern remained unchanged. Thus, in comparing proteins presumed to be closely related, that is, ox, pig, rabbit, and chicken insulins, the chromatographic behavior of one hormone tells

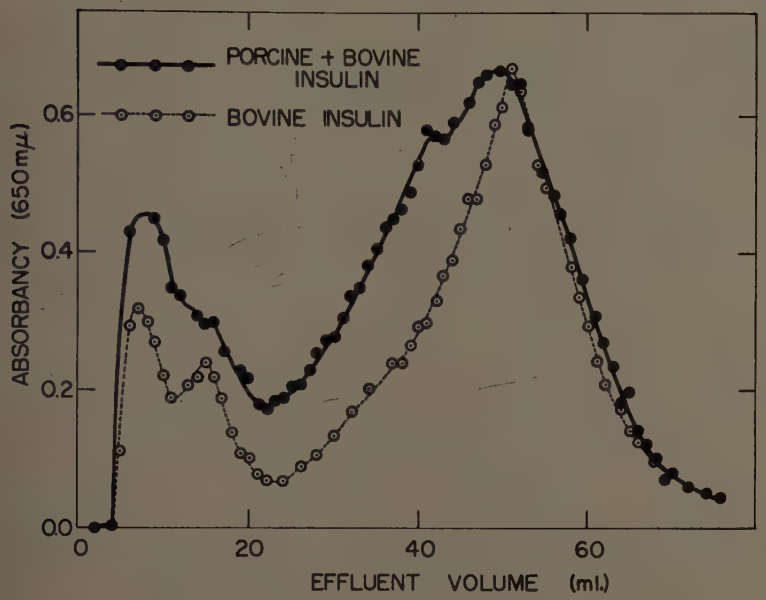


FIGURE 18. A chromatogram of crystalline beef insulin compared to a chromatogram of crystalline beef insulin mixed with crystalline pork insulin. The chromatography was performed on 0.64 × 30 cm. columns of Amberlite IRC-50 with 0.13 *M* phosphate-5 *M* urea buffers of *p*H 6.0.

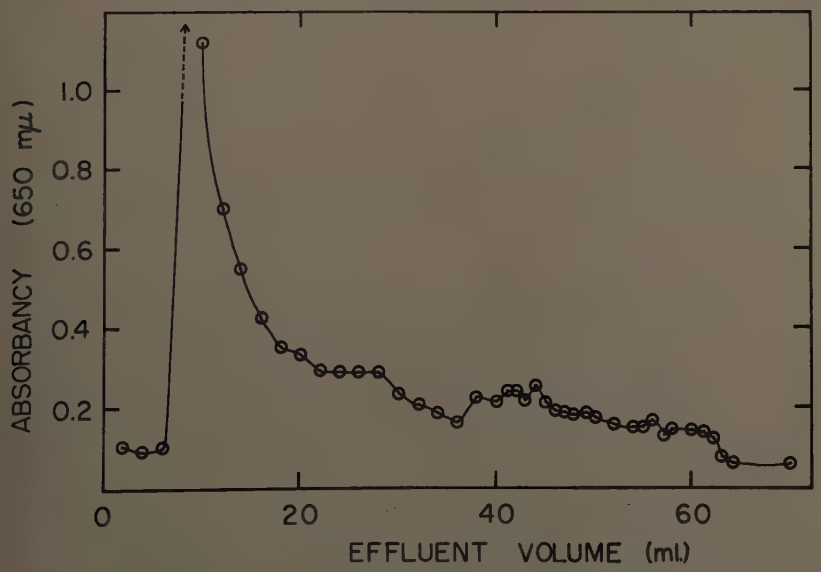


FIGURE 19. The chromatography of crude chicken insulin (1 U./mg.) on a 0.64 × 30 cm. column of Amberlite IRC-50 using a *p*H 6.0 buffer of 0.13 *M* phosphate-7 *M* urea.

little or nothing about the behavior of another, and at least part of the variability is due to a different sensitivity of the proteins to urea.

Another example of the difficulty in predicting relative elution volumes of proteins is afforded by the ion-exchange chromatography of glucagon. While the existence of this pancreatic hyperglycemic hormone was postulated by Kimball and Murlin in 1923,[19] it was only recently that glucagon was isolated by Staub *et al.*[20] and shown to be a basic polypeptide.

Since insulin had already been chromatographed on Amberlite IRC-50 using 0.13 *M* phosphate-7 *M* urea buffer at *p*H 6.0, these were the first conditions

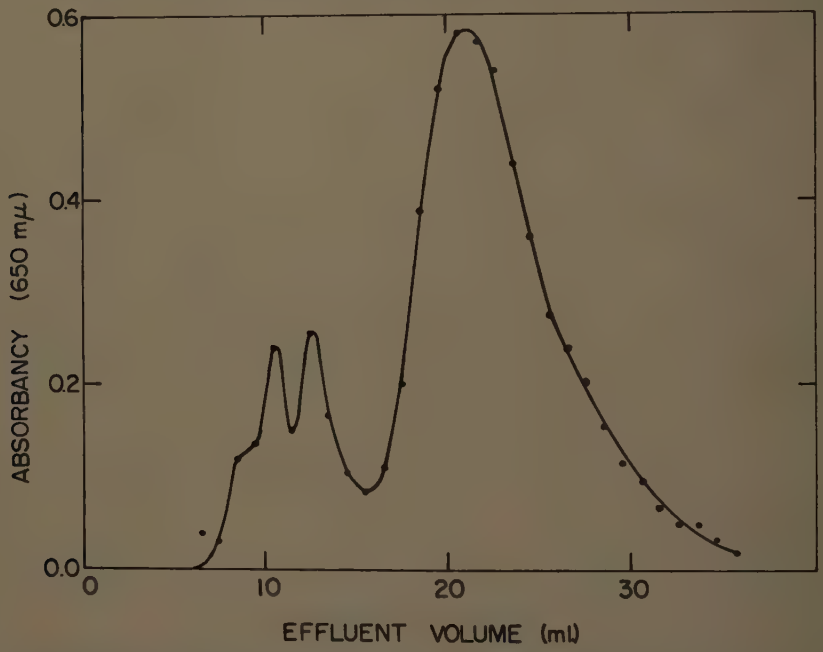


FIGURE 20. The chromatography of glucagon on Amberlite IRC-50 (0.64 × 30 cm.) in 0.13 *M* phosphate-7 *M* urea at *p*H 6.0.

tried for glucagon in the hope of determining the position of the latter hormone on the chromatograms of insulin, but these conditions were tried in spite of the fact that glucagon has a more basic isoelectric point than insulin and would therefore be expected, on the basis of simple ion exchange, to be excessively retained under conditions appropriate for insulin. Chromatography in 0.13 *M* phosphate-7 *M* urea gave the results shown in FIGURE 20 in which it may be observed that the main component of glucagon is eluted at 21 ml. This elution volume is slightly less than the 30 ml. elution volume of insulin rather than much greater. To make certain that the unexpected behavior of glucagon was not due to transformation, the main component was assayed and found to be active; and the recovery of peptide material, based on the yield of color, was checked and found to be quantitative (97 per cent). A further test

was made by reapplying to the column eluate taken from a previous chromatogram. The results of such an experiment are shown in FIGURE 21, which reveals a single peak at the appropriate elution volume. Furthermore, a sample of glucagon was submitted to chromatography after prior incubation in the phosphate-urea buffer for eight hours. The results obtained showed no perceptible change from the earlier chromatographic pattern.

The decreased retention of glucagon relative to that of insulin thus seems to be real. The fact that the order of elution of glucagon and insulin is contrary to expectation may be due to the greater unfolding of the insulin molecule in

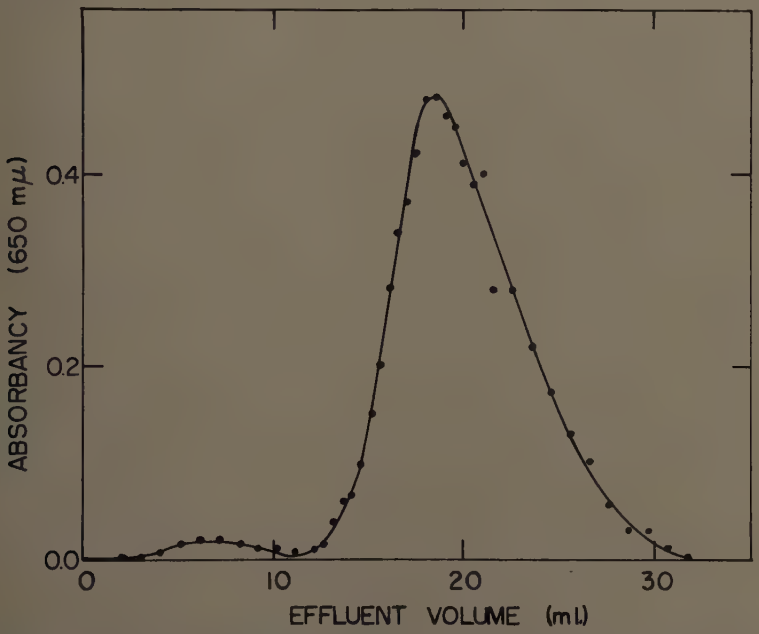


FIGURE 21. The rechromatography of eluate from the main peak of a previous chromatogram run in 0.13 *M* phosphate-7 *M* urea at *p*H 6.0 on a column of Amberlite IRC-50, 0.64 × 30 cm.

urea, but this seems unlikely in such small and relatively simple molecules as insulin and glucagon. Furthermore the data in TABLE 1 indicate that insulin and glucagon have about the same sensitivity to urea (in this range of urea concentration).

It might be suggested, then, that the relatively greater retention of insulin is due simply to its larger size. This notion might be tested by the study of a protein that is similar to insulin except for being larger. Such a protein is prolactin. The isoelectric points, the solubility patterns, and the stabilities of insulin and prolactin are very similar. A mixture of the two proteins would be exceedingly difficult if not impossible to separate by usual means. The main chemical difference between insulin and the pituitary hormone is that the latter has a molecular weight of 25,000.[21] When prolactin was submitted

to chromatography under conditions similar to those used for insulin, using 8 *M* urea, the pattern shown in FIGURE 22 was observed. This hormone has already been shown to contain three active components when examined by

TABLE 1
THE RETENTION VOLUME OF GLUCAGON AND INSULIN AT DIFFERENT CONCENTRATIONS OF UREA

Urea concentration	Retention volume* (ml.)	
	Glucagon	Insulin
8 *M*	2	6
7 *M*	13	22
6 *M*	22	32

* The retention volume used here represents the total elution volume minus the hold-up volume (8 ml.) of the column. Chromatography was performed using 0.64 × 30 cm. columns of Amberlite IRC-50 with urea-containing buffers of 0.13 *M* phosphate at *p*H 6.0.

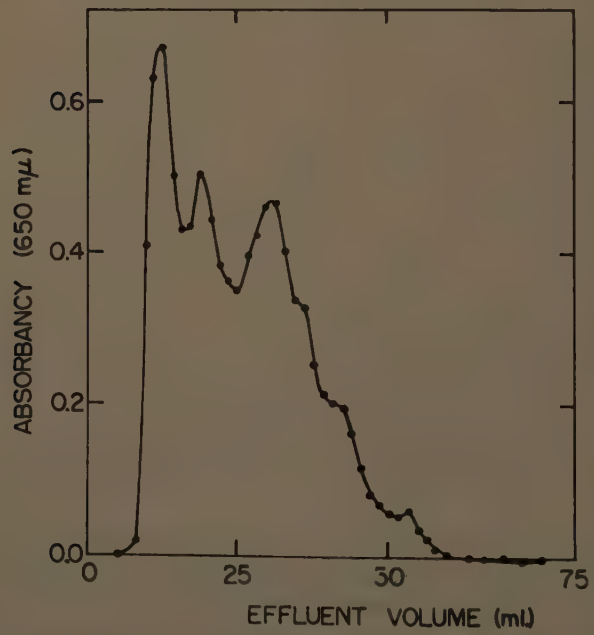


FIGURE 22. The chromatography of prolactin in 0.13 *M* phosphate-8 *M* urea buffer at *p*H 6.0 on a 0.64 × 30 cm. column of Amberlite IRC-50. All three peaks were fully active.

electrophoresis[22] or by counter-current distribution;[23] therefore the complexity of this chromatogram was not a surprise. The significance of the pattern for the present discussion is that the elution volumes of the prolactins are quite comparable to those of insulin even though prolactin is much larger.

In 8 *M* urea, then, the much greater size of prolactin relative to insulin would not seem to determine the elution volume in any profound way. Sig-

nificantly, however, the elution volume of prolactin is much more sensitive to changes in urea concentration than is the elution volume of insulin. The elution volume of prolactin is so drastically increased by dropping the urea concentration from 8 *M* to 6 *M* that most of the prolactin is still not eluted after 100 to 150 ml. when 6 *M* urea is used. These findings suggest that in 8 *M* urea most of the resin-protein hydrogen bonds are eliminated and ion exchange is playing the larger role, but as the urea concentration is lowered increasing numbers of hydrogen bonds form between resin and protein. Of course, the larger the protein, the sooner it will form a large enough number of hydrogen bonds to be termed multipoint attachments.

By operating in the region in which both ion exchange and hydrogen bonds are involved, the sensitivity of the chromatographic system may be greatly increased, as was seen in the case of the insulins. It will be of interest to see just how subtle the differences are between pork and beef insulin on the one hand and rabbit insulin on the other; especially as rabbit insulin is chromatographically intermediate between desamido beef insulin and pork insulin, which differ from the main component of beef insulin respectively by the loss of an amide group and the exchange of a couple of amino acids.

It will also be of interest to see if the greater sensitivity of the chromatography of larger proteins to urea is a general phenomenon. It would certainly be pleasing to have this extra parameter to manipulate in the chromatographic fractionation of proteins that differ mainly in shape and size, as do prolactin and insulin.

The use of buffers containing such high concentrations of urea would seem an unlikely way to achieve a completely general method of chromatography of proteins. It is to be hoped, however, that this technique will extend the use of resin chromatography to several other proteins and protein derivatives.* Furthermore it may be that the disruption of hydrogen bonding by lesser concentrations of urea would improve the yields of protein obtained by other methods of chromatography, since the hydrogen bonding phenomenon in some degree is a general problem in chromatography affecting yields and resolution.

Acknowledgments

These experiments were begun at The Rockefeller Institute, New York, N. Y., in the laboratories of Stanford Moore and William H. Stein, to whom we are grateful for many valuable discussions. We are grateful to Edward Grinnan of Eli Lilly and Co., Indianapolis, Ind., for supplying most of the insulin samples and the glucagon samples (page 550). The fractions of insulin purified by counter-current distribution were generously given by Lyman Craig of The Rockefeller Institute (page 553). In the peak referred to on page 553, without correction for moisture or ash, it was found to contain 21 U./mg.; we are grateful to Edward Grinnan of Eli Lilly and Co. for having this assay performed. Andreas Chrambach and Frederick Carpenter of the University of California, Berkeley, Calif., generously furnished the samples of insulin puri-

* One impetus for the initiation of this work was the desire to chromatograph protein derivatives 24, 25 which like insulin, glucagon, and prolactin were insoluble in neutral aqueous solutions.

fied by partition chromatography (page 557). The assay referred to on page 566 was kindly performed by William Bromer of Eli Lilly and Co., to whom we are greatly indebted.

References

1. Moore, S. & W. H. Stein. 1956. Advances in Protein Chem. **11:** 191.
2. Schmid, K., M. B. MacNair & A. F. Burgi. 1958. J. Biol. Chem. **230:** 853.
3. Boardman, N. K. & S. M. Partridge. 1955. Biochem. J. **59:** 543.
4. Boardman, N. K. 1959. J. Chromatography. **2:** 398.
5. Sober, H. A. & E. A. Peterson. 1954. J. Am. Chem. Soc. **76:** 1711.
6. Best, C. H., R. E. Haist & J. H. Ridout. 1940. J. Physiol. **97:** 107.
7. Cole, R. D. & C. H. Li. 1955. J. Biol. Chem. **213:** 197.
8. Somogyi, M. 1945. J. Biol. Chem. **160:** 69.
9. Riddle, O., R. W. Bates & S. W. Dykshorn. 1933. Am. J. Physiol. **105:** 191.
10. Hirs, C. H. W., S. Moore & W. H. Stein. 1953. J. Biol. Chem. **200:** 493.
11. Lowry, O. H., N. J. Rosebrough, A. L. Farr & R. J. Randall. 1951. J. Biol. Chem. **193:** 265.
12. Harfenist, E. J. & L. C. Craig. 1952. J. Am. Chem. Soc. **74:** 3083.
13. Harfenist, E. J. 1953. J. Am Chem. Soc. **75:** 5528.
14. Bischoff, F. & A. K. Bakhtiar. 1956. J. Am. Chem. Soc. **78:** 1343.
15. Carpenter, F. H. & G. P. Hess. 1956. J. Am. Chem. Soc. **78:** 3351.
16. Chrambach, A. C. 1959. Doctoral dissertation. Univ. Calif. Berkeley, Calif.
17. Samsonov, G. V. & M. D. Faddeeva. 1956. Biokhimiya. **21:** 403.
18. Kupke, D. W. & K. Linderstrøm-Lang. 1954. Biochim. et Biophys. Acta. **13:** 153.
19. Kimball, C. P. & J. R. Murlin. 1923–1924. J. Biol. Chem. **58:** 337.
20. Staub, A., L. Sinn & O. K. Behrens. 1955. J. Biol. Chem. **214:** 619.
21. Li, C. H., R. D. Cole & M. Coval. 1957. J. Biol. Chem. **229:** 153.
22. Cole, R. D. & C. H. Li. 1959. Biochim. et Biophys. Acta. **33:** 563.
23. Cole, R. D. & C. H. Li. 1958. Arch. Biol. Biophys. **78:** 392.
24. Bailey, J. L. & R. D. Cole. 1959. J. Biol. Chem. **23:** 1733.
25. Moore, S., R. D. Cole, H. G. Gundlach & W. H. Stein. 1958. Symposium on Proteins, 4th Intern. Congr. Biochemistry. : 52. Vienna, Austria.

HOMOGENEITY STUDIES WITH INSULIN AND RELATED SUBSTANCES

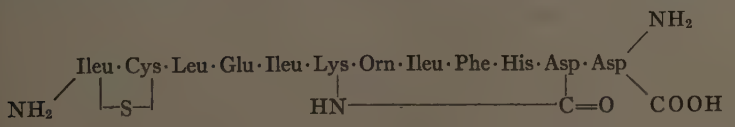
Lyman C. Craig, Te Piao King, William Konigsberg
The Rockefeller Institute, New York, N. Y.

It is obvious that any comprehensive investigation of the detailed structural features of complicated molecules must include an inquiry into the limitations of our ability to separate individual substances and recognize whether and to what degree we have separated them. With proteins this leads us to question the degree of uniqueness we might expect with regard to amino acid composition, sequence, and conformation. The latter includes so-called denaturation phenomena exclusive of those that derive from a change in the primary structure of the molecule. It is only to be expected that such large molecules would present almost limitless opportunities for subtle changes in shape. A further complicating property is that of association. It seems that proteins more often than not prefer to associate at least partly, or to bind other solutes, including other proteins.

In our laboratory we are investigating these problems with a variety of tools, but those being emphasized most are countercurrent distribution (C.C.D.), membrane diffusion,[1,2] and the ultracentrifuge. The first depends on certain properties of the molecule that determine relative polarity or the balance between hydrophobic and hydrophylic properties. The second, like the third, arises from and permits assessment of molecular size and/or shape. It is of primary interest to ask how selective these methods really are in terms of the structural differences they can reveal.

To answer this question most conclusively, models with known structures should be available. The difficulty here is that we do not have models of molecular weight much above 1000 for which the conformation as well as the sequential formula is really known. Nevertheless, this paper will report a number of experiences with certain of the best available models: the bacitracins, the insulins, lysozyme, and the ribonucleases.

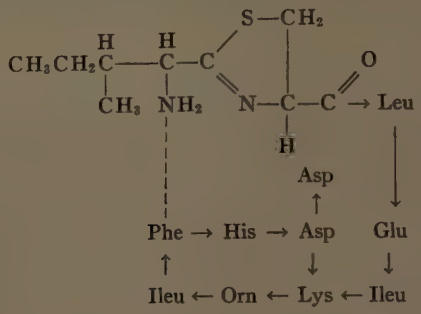
Structural studies with bacitracin A[3,4] have led to FORMULA I as the most probable arrangement. It is a dodecapeptide of molecular weight 1422 that has become tangled with itself in such a way that at least two covalently bonded



FORMULA I. Amino acid sequence in bacitracin.

rings are involved (FORMULA II). We have found that by countercurrent distribution we can separate easily another bacitracin, called B, that differs from A by having an additional valine residue. As improvement in separating power developed, it was found that the earlier preparations of neither A nor B were really pure but were mixtures of isomers. A could be separated[5] by prolonged countercurrent distribution or cellulose ion-exchange chromatog-

raphy into two isomers that differed only in the optical configuration of an α-carbon atom of the *N*-terminal isoleucine. This carbon is able to mutaro-



FORMULA II. Ring structure in bacitracin.

tate. An example is thus provided of the way the configuration of an amino acid residue can be changed by the induction effects of the groups around it.

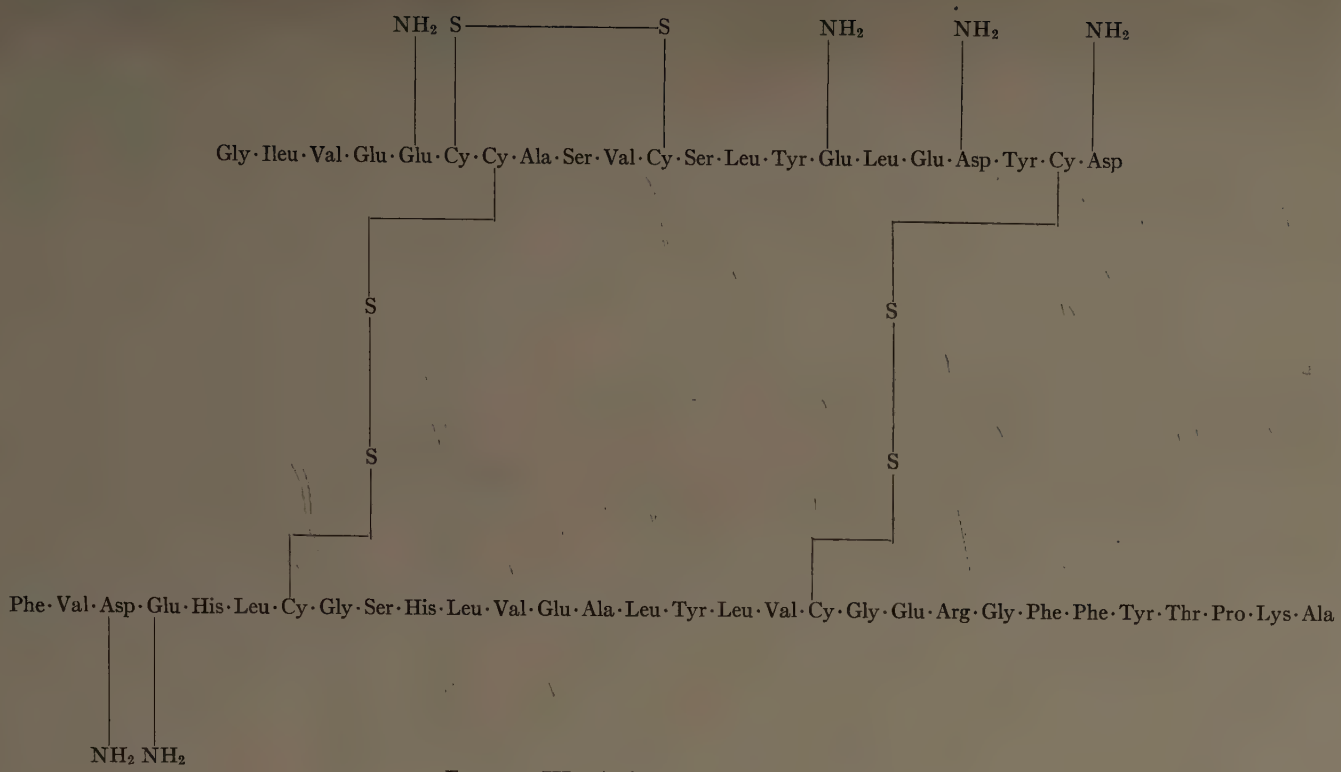
It was found that B not only was a mixture arising from such a mutarotation, but also was a mixture of sequential isomers, one with an *N*-terminal isoleucine and the other with an *N*-terminal valine (unpublished observations). Mutarotation during the distribution prevented their complete separation. We cannot say from this experience that we would never miss impurity due to structural differences of this magnitude, but it does give something to use as a basis for speculation.

Let us now pass to insulin, a molecule roughly four times the size of bacitracin. Beef insulin has the sequential formula shown in FORMULA III, according to Ryle *et al.*[6] It has 51 amino acid residues and a molecular weight of 5733. The earlier attempts to study the purity of a highly purified sample of beef insulin by countercurrent distribution[7] gave the pattern shown in FIGURE 1.* E. J. Harfenist, who did this work, found the smaller peak to have one less amide group, or a total of five, compared to the larger band with six. They were of equal biological activity and had the same amino acid content. We found we could easily separate mono- and di-DNP derivatives from untreated insulin and, in this way, we showed that the minimum molecular weight was not 12,000, as had been thought, but actually was in the 6000 range.[8]

Later, after Harfenist had shown the differences in the quantitative amino acid composition of beef, pork, and sheep insulins,[9] we set up an experiment with an artificial mixture of beef and pork insulins to see if they could be separated. The distribution pattern, FIGURE 2, indicated this was possible. In so far as we know,[10] the structural difference between these two insulins is the substitution of a threonine for an alanine and an isoleucine for a valine.

Studies with a larger molecule, ribonuclease A, can now be considered. It has a molecular weight of 13,600. The amino acid sequence of the 124 residues is discussed by C. H. Hirs elsewhere in this monograph. Fractionation studies of a crystalline preparation by ion-exchange chromatography[11] had shown the presence of small percentages of other components.

* Lot T-2344, Eli Lilly and Company, Indianapolis, Ind.



FORMULA III. Amino acid sequence in insulin.

In searching for a satisfactory system for C.C.D., T. P. King[12] found that denaturation effects interfered too much except with an ammonium sulfate-ethanol system. It gave patterns as shown in FIGURE 3.* The separate fractions were cross-checked by ion-exchange chromatography as shown in the Figure. The main component amounted to only about 70 per cent of the total.

A different sample† had been shown by ion-exchange chromatography[11] to have more of a different component. By C.C.D., D. Eaker and King (unpublished observations) obtained the pattern shown in FIGURE 4. Again, each

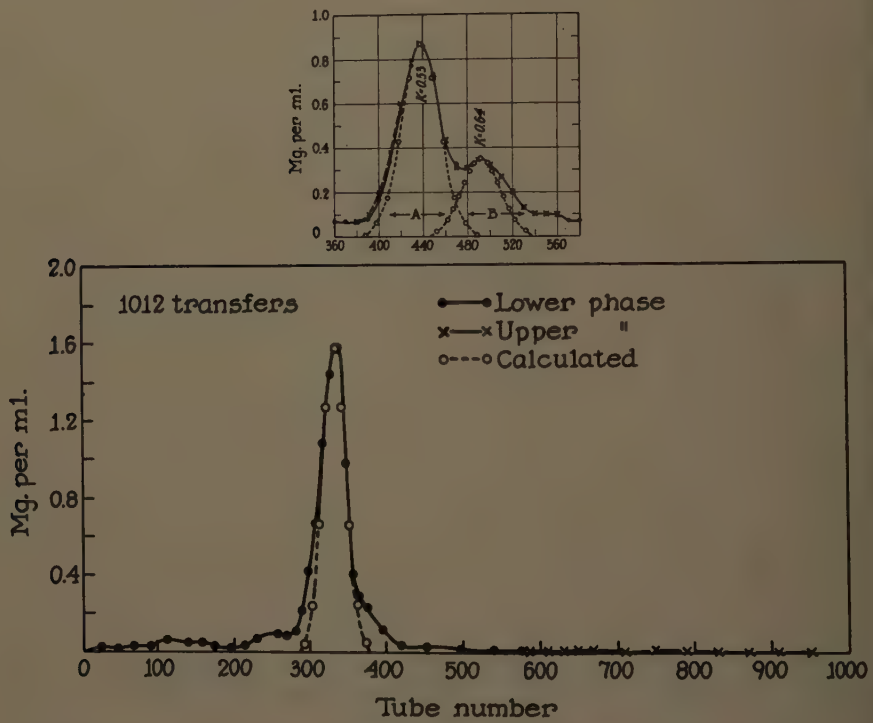


FIGURE 1. C.C.D. patterns with insulin (see text). The upper pattern is with the original. The lower is with the IUPAC sample at 1000 transfers.

fraction was cross-checked by ion-exchange chromatography. We shall not discuss the details of this separation but shall speak of only one component, the one to the right of the main component in this distribution pattern. It was found to have the same enzymatic activity as A, but amino acid analysis and end-group determination showed it to differ from A by the lack of the *N*-terminal lysine. The structural differences inherent in the other components, revealed by C.C.D. and ion-exchange chromatography, are not yet known, but probably they are of the same magnitude as the one already demonstrated. It is of interest to note the close correspondence of the theoretical curve to the experimental one.

* Armour Lot 381-059.
† Armour Lot 381-062.

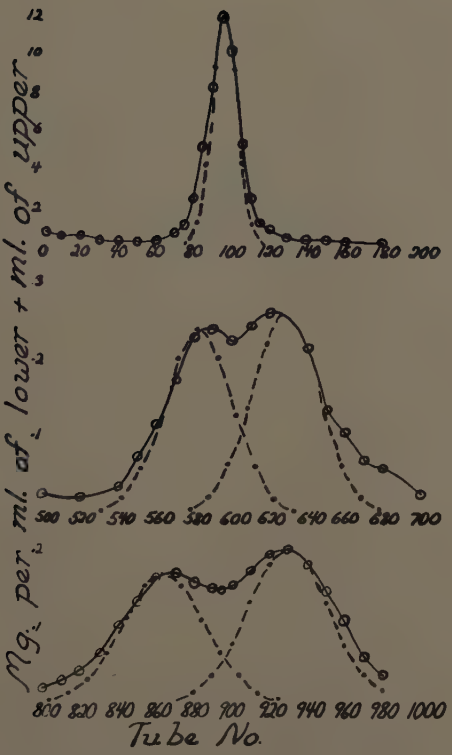


FIGURE 2. Separation patterns of an artificial mixture of beef and pork insulins by C.C.D. at 200, 1264, and 1850 transfers (pyridine system).

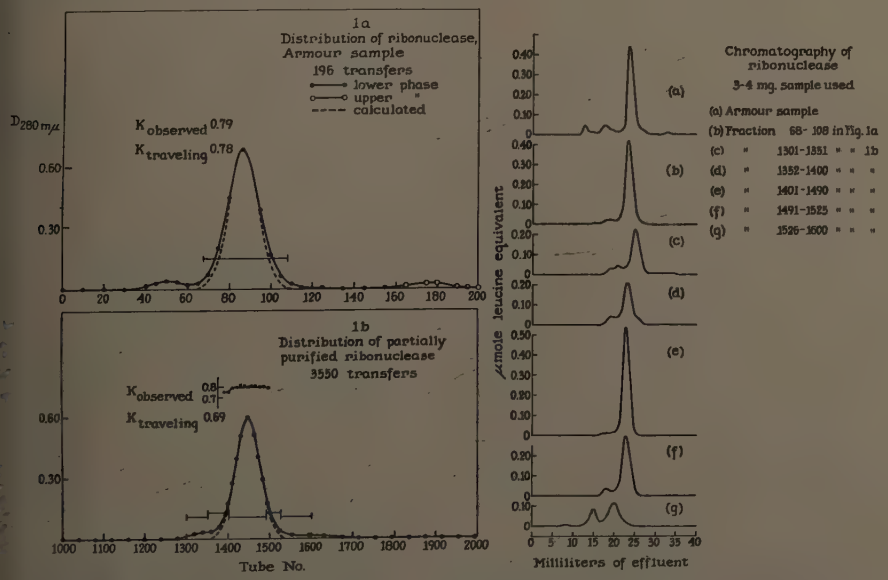


FIGURE 3. Patterns of C.C.D. and ion-exchange chromatography with ribonuclease (see text).

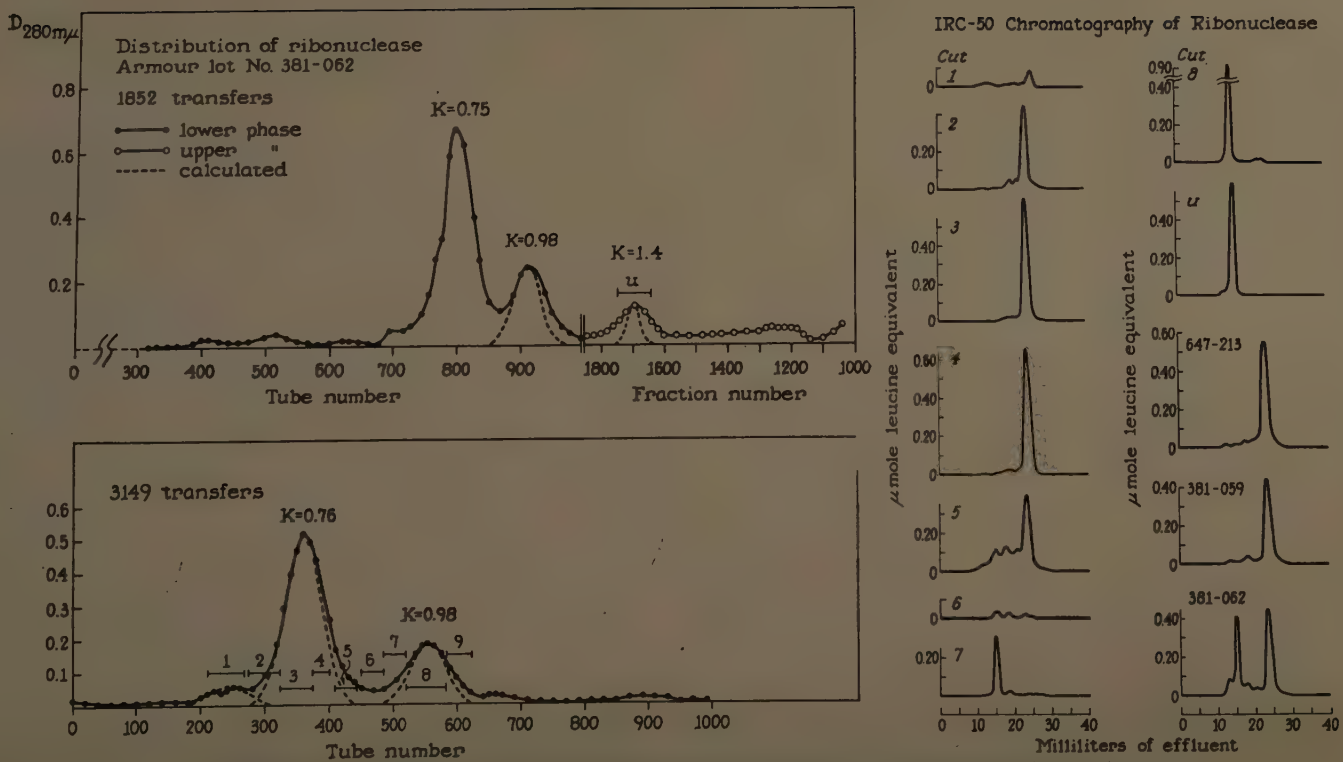


FIGURE 4. Patterns of C.C.D. and ion-exchange chromatography with ribonuclease (see text).

C.C.D. studies with lysozyme (mol. wt. 14,800) in a similar system[12] were even more clear cut as may be seen from FIGURE 5.*

With this short account of our experience with molecules a little more than double the size of insulin, let us return to a closer examination of the status of its purity. Several years ago we were asked to study the purity of a sample of insulin to be given out as a standard sample under the auspices of the protein committee of International Union of Pure and Applied Chemistry (IUPAC).†

Using the original 1 per cent dichloroacetic acid system,[7] we obtained the lower pattern shown in FIGURE 1. The upper pattern is that obtained in the

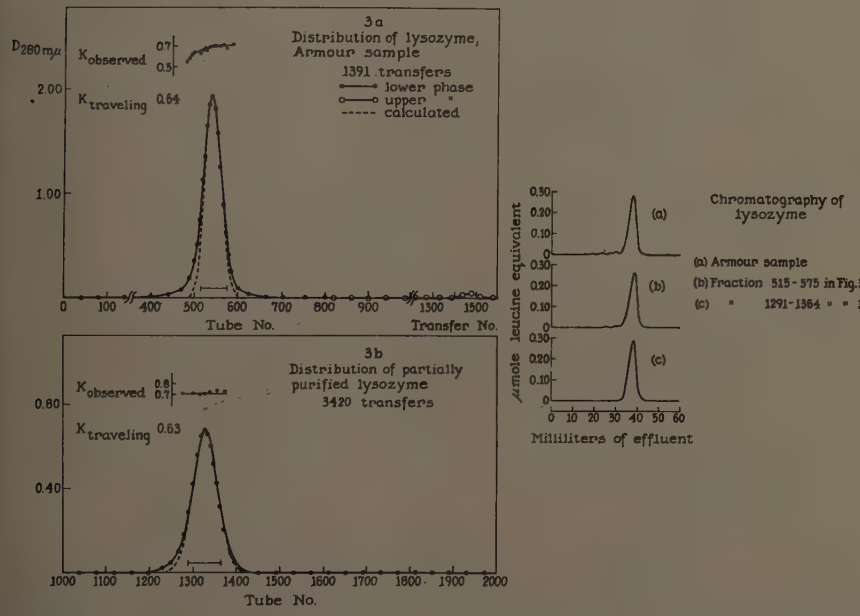


FIGURE 5. Patterns of C.C.D. and ion-exchange chromatography with lysozyme (see text).

earliest work where the desamido derivative was isolated. The IUPAC sample is obviously low in the desamido derivative. Several smaller components have been thrown off.

At this point another insulin system was developed, made from 0.1 per cent aqueous acetic acid, *N*-butanol, and pyridine (volume proportions: 11, 5, 3). It is of interest to recall that in the original crystallization of insulin[13] a mixture of aqueous pyridine and acetic acid was used as a solvent. Loss of activity was not noted.

A question concerning C.C.D. under investigation at the time of our pyridine acetate study was that of the speed at which equilibrium could be established on each transfer. FIGURE 6 is presented, in part, from the standpoint of these

* Armour Lot 381-187.
† Batch 2189.

studies. The two patterns were obtained in the same system, with the same machine, and with the same IUPAC sample.

For the top pattern only 5 relatively rapid tips requiring a total of 15 sec. were given on each transfer. In the case of the lower pattern the rate of tipping was slowed to one third, 5 tips in 45 sec. Here 2 cuts were taken for

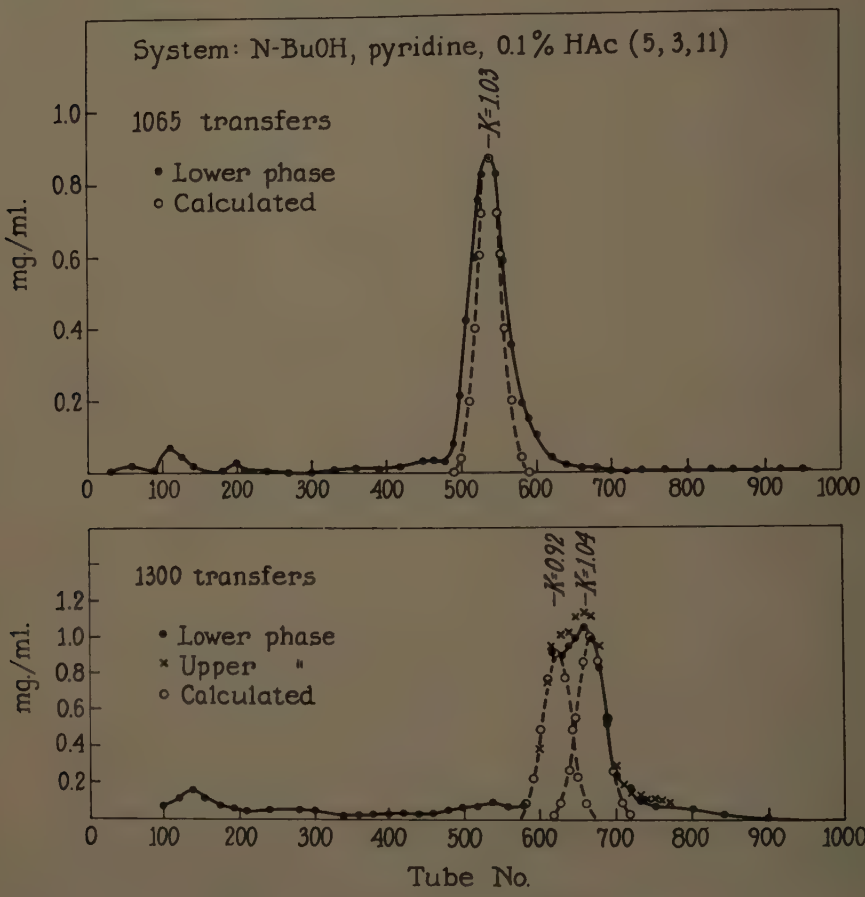


FIGURE. 6. C.C.D. patterns with IUPAC insulin.

recovery. No difference in amino acid composition or rotatory dispersion could be found. Inhomogeneity due to loss of an amide did not seem likely because distribution in the dichloroacetic acid system, FIGURE 1, had revealed not more than a few per cent of the desamido form. It was suspected that the differences were due to some type of conformational isomerism, a difficult theory to prove conclusively.

In a sedimentation experiment with 30 per cent acetic acid in the ultracentrifuge, D. A. Yphantis found a difference. The cut on the left in an equilibrium run behaved as if it had an average mol. wt. of 7400, while that on the right ap-

peared to be 6200. If this difference is real, it would indicate that the cut on the left does not dissociate into the monomeric insulin as completely as the one on the right. This could imply a conformational difference.

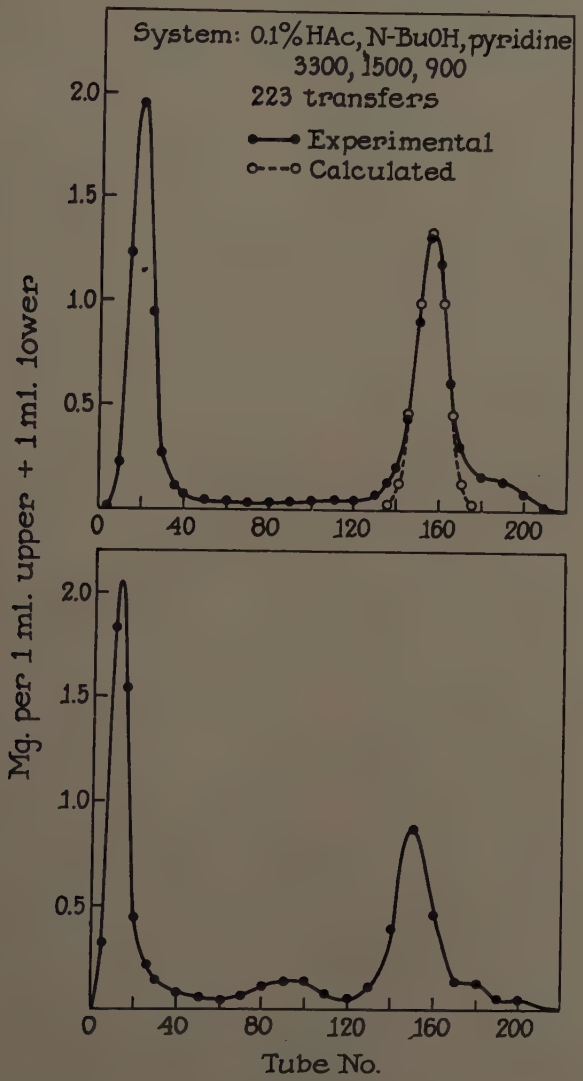


FIGURE 7. C.C.D. patterns with oxidized insulin; upper with one sample (see text), lower with IUPAC sample.

The homogeneity question was considered also, in another way. It had been found that one sample of insulin gave a very clean oxidation to the A and B chains with performic acid at 0° C.[14] The distribution pattern is shown in FIGURE 7.* The IUPAC sample, however, did not give as clear a result, as can

* Lilly Lot 200-1B-11C.

be seen from the lower pattern. More solute was present in the region apart from the two main bands, and the yield of B, the band on the right, was low.

The sample that gave the upper pattern was investigated further by C.C.D. in the pyridine system. The pattern obtained, FIGURE 8, was definitely different from that obtained with the IUPAC sample. Another difference was noted. About 15 per cent of the sample refused to dissolve in the system, while essentially all the IUPAC sample dissolved.

Activity determinations* on recovered material indicated the peak material to be fully active. Cuts taken from other regions, including the insoluble material seemed somewhat less active.

Still another research sample of crystalline beef insulin† was distributed in the pyridine system by the slow tipping. At 1000 transfers it gave the pattern shown in FIGURE 9. Aside from a slight turbidity, it was completely soluble in the system, at least in the concentration level employed.

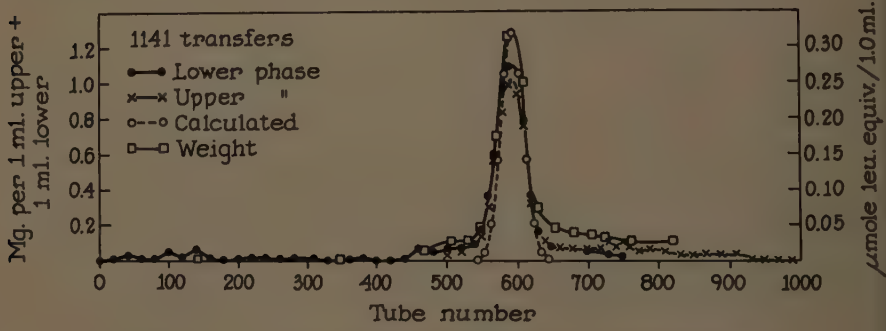


FIGURE 8. C.C.D. pattern of insulin (see text) in pyridine system.

We now wondered if the 2 bands could be recovered and redistributed with retention of their individuality. Accordingly the cuts shown were recovered and each redistributed to 1000 transfers, as shown in the lower patterns. In each case, only a single band was obtained although it appeared to be definitely wider than the theoretical. Each band was recovered in 2 approximately equal cuts. Careful determination of partition ratios indicated that the slowest moving material had a K of 1.00, while the most rapidly moving cut had a K of 1.18.

These experiments definitely show the presence of at least two major components. To the best of our information they are isomeric, and the possibility must be considered that they could be interconvertible under the right conditions. Indeed, such a partial conversion might explain the spreading of the bands noted above.

The complications inherent in the study of the purity of insulin are particularly emphasized by its strong tendency to associate. It is known from extensive ultracentrifugal studies[15] that only the strongest dissociating solvents and low *p*H, such as 30 per cent acetic acid, are capable of making it behave

* Made by the Eli Lilly Company.
† Armour Lot 784-139-19.

as a monomer. Purity studies with different preparations of insulin[16] by the ultracentrifuge have shown much greater differences at higher *p*H values than at low. Here the question must be asked whether the difference is due to states of aggregation of not readily reversible forms perhaps resulting in part from the possible requirement of a shift in conformation of the monomer. It would be of great interest to know to what extent shifts in conformation in the

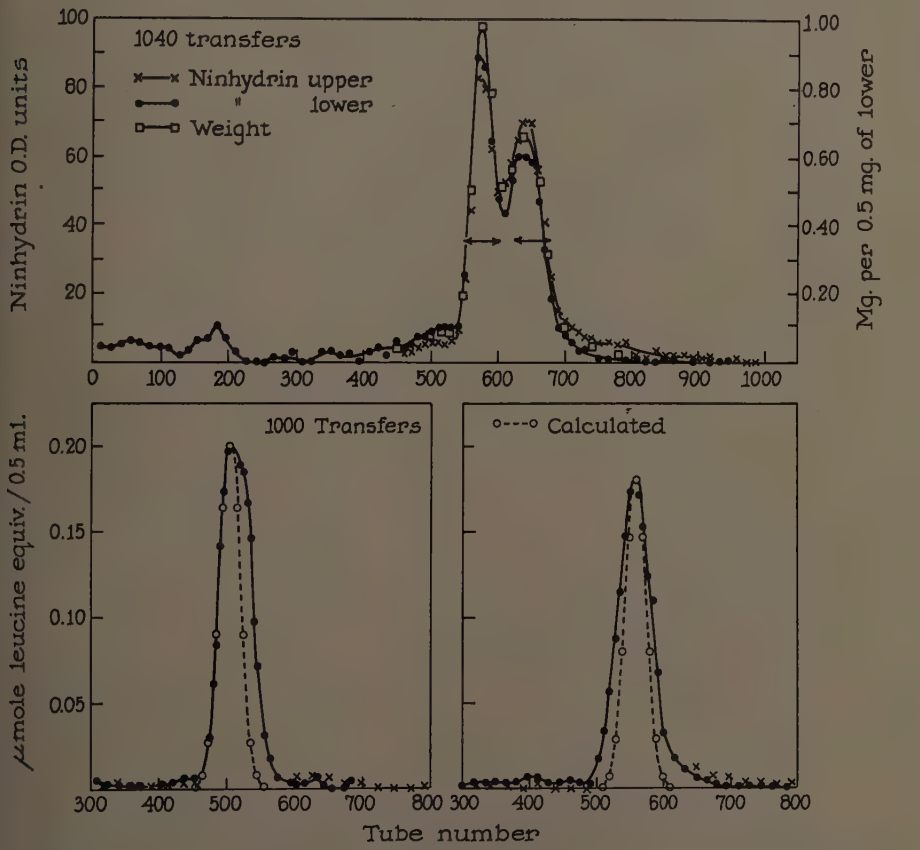


FIGURE 9. C.C.D. patterns of insulin (see text). The two lower patterns are redistributions of the respective bands.

monomer are related to the polymerization phenomenon. They may be different in those forms ranging from the dimer to the insoluble fibril.[17]

It is known that, with the extreme forms, the fibrils form at a measurable rate, apparently requiring definite patterns or templates in the form of seeds.[17] Indeed, the C.C.D. system containing dichloroacetic acid and used in FIGURE 1 for separation of the desamido form is useless if seeds are present. All the insulin will precipitate in the machine. Reconversion to the soluble form takes place only under the strain of strong alkali.

Information concerning the rate of polymerization or change of shape with

the smaller more soluble forms can be obtained from membrane diffusion studies.[2] In this method the rate of diffusion through a cellophane membrane of suitable porosity is measured and the result expressed as a semilogarithmic plot of per cent of the original remaining inside the membrane against time. The greatest selectivity has been obtained when the solute of interest would barely pass, a limitation requiring the use of a maximum dialyzing area in the diffusion cell.

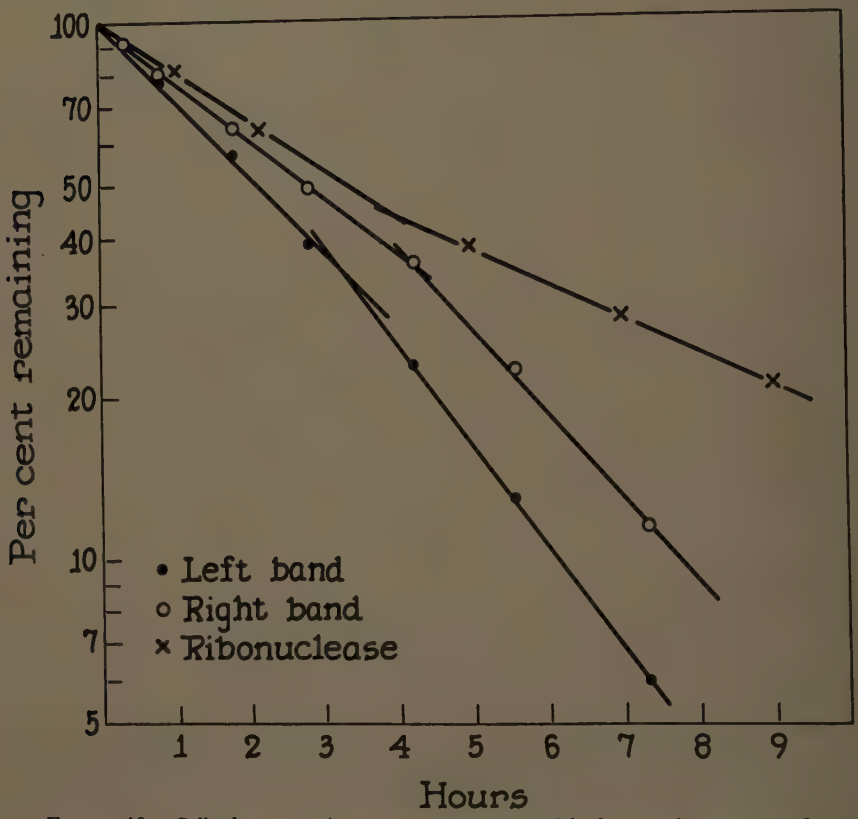


FIGURE 10. Cellophane membrane escape patterns with the cuts from FIGURE 9.

A sample of Lilly beef insulin further purified by C.C.D. in the original DCA system[7] gave the lower escape curve shown in FIGURE 10 with 0.1 *M* acetic acid as the solvent. It would appear that this sample is reasonably homogeneous with respect to size. Almost certainly, it is present in this solution, at least partly, as the monomer since its rate of diffusion was nearly double that of ribonuclease (mol. wt. 13,800). The IUPAC sample that gave the 2 bands on distribution in the pyridine system gave a membrane diffusion result not different from that of the purified sample.

On the other hand, a Lilly sample of amorphous insulin gave the dashed curve of FIGURE 11 showing the presence of considerable amounts of polymers

or molecules of less compact form in the solution. When the diffusion of crystalline beef insulin was done at a lower *p*H, the rate became slower. This was also the case when the diffusion was done in the presence of urea, which is to be expected from the currently accepted theory of the effect of urea on a

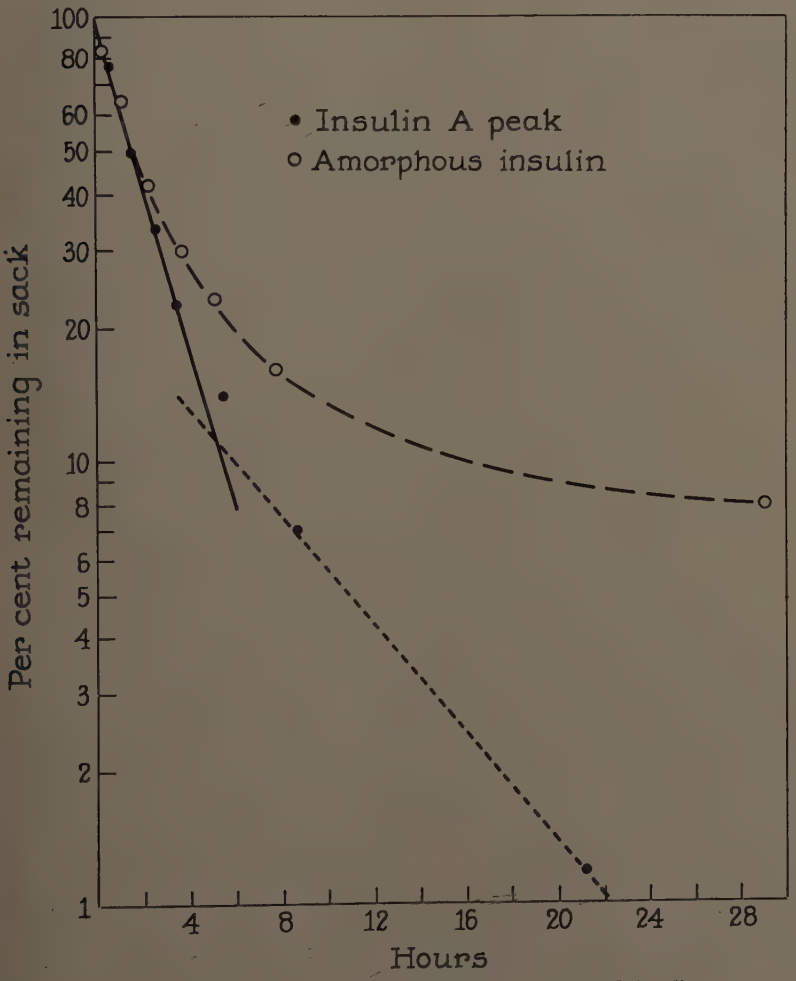


FIGURE 11. Cellophane membrane escape patterns with insulins.

protein. Almost any transformation, such as performic acid oxidation, results in a slower rate of diffusion except where smaller fragments are formed, as in the case of insulin.

A diffusion experiment with insulin in the membrane used for FIGURE 10 was performed by suspending the sample in water and adding just sufficient ammonia to cause solution. The *p*H at this point was about 8. Dialysis against distilled water did not cause any of the insulin to pass through the

membrane, although the *p*H dropped to about 6. This is in accord with expectations from the ultracentrifuge data that have indicated a molecular size at this *p*H in the range of 36,000 to 48,000 or higher.

If this is the case, an investigation of the molecular size of insulin in the pyridine acetate, *N*-butanol system used for the fractionation studies is of interest. The *p*H of this system approximates 6.7. It was found that such diffusion experiments could be done in the aqueous phase of the system and that the majority of the sample would diffuse through at a constant rate, suggesting a monomer or at most a dimer. Here, too close an assessment cannot be made because the butanol and pyridine themselves have a strong retarding effect on the rate of dialysis. The rate of dialysis was found to be concentration-dependent. These conclusions were checked by Yphantis in the ultracentrifuge in essentially the lower phase of the system to which sodium chloride had been added to a 0.15 *M* concentration.

It was somewhat disappointing to find that although small differences between the various insulins could be found by diffusion in the pyridine acetate, these differences could not be correlated clearly with the C.C.D. experiments. However, they do indicate that in the C.C.D. experiments the separation is largely concerned with monomers and dimers.

When the two fractions obtained from the distributions shown in FIGURE 9 were compared in 0.01 *N* acetic acid by the membrane diffusion technique, a very interesting result was obtained, FIGURE 10. A definite difference in their diffusion rate was found. Moreover, both showed a change in escape rate as the concentration decreased that indicated an increased rate of diffusion with dilution. This is the type of escape pattern to be expected from a dissociating dimer. The escape curve for ribonuclease in the same membrane is given for comparison. Here, as with many proteins studied,[2] a change in escape rate is noted that would indicate a dimer (or change in conformation) whose rate of dissociation is slow or not reversible under these conditions.

When absorption spectrum measurements[1] were made on each of the successive diffusates, a further interesting observation was made. The first diffusate contained pyridine, as might be expected since the two fractions had been recovered from a pyridine solution by direct lyophilization. The other diffusates gave the type of absorption expected from insulin, but with a lowered optical density per milligram of residue in the region of the tyrosine absorption for the two last diffusates as compared to the second and third. It has been known for a long time that the quantitative aspects of the ultraviolet absorption spectra would vary depending on the way the sample was prepared.[16] Recently, the careful studies of Laskowski *et al.*[18] have connected this to the hydrogen bonding of the tyrosines. The more highly associated forms give a higher molecular absorption. Conversely, the dissociation from the dimeric to the monomeric form gives a lowered molecular absorption, as was noted in our experiments.

Acknowledgment

We thank H. Jaffe of The Rockefeller Institute for the absorption spectrum measurements.

References

1. Craig, L. C., T. P. King & A. Stracher. 1957. J. Am. Chem. Soc. **79:** 3729.
2. Craig, L. C., W. Konigsberg, A. Stracher & T. P. King. 1958. Symposium on Protein Structure. : 104. A. Neuberger, Ed. John Wiley & Sons, Inc. New York, N. Y.
3. Craig, L. C. & W. Konigsberg. 1957. J. Org. Chem. **22:** 1345.
4. Swallow, D. L. & E. P. Abraham. 1959. Biochem. J. **72:** 326.
5. Konigsberg, W. & L. C. Craig. 1959. J. Am. Chem. Soc. **81:** 3452.
6. Ryle, A. P., F. Songer, L. F. Smith & R. Kitai. 1955. Biochem. J. **60:** 541.
7. Harfenist, E. J. & L. C. Craig. 1952*a*. J. Am. Chem. Soc. **74:** 3083.
8. Harfenist, E. J. & L. C. Craig. 1952*b*. *Ibid.* **74:** 3087.
9. Harfenist, E. J. 1953. J. Am. Chem. Soc. **75:** 5528.
10. Sanger, F. 1959. Science. **129:** 1340.
11. Hirs, C. H. W., S. Moore & W. H. Stein. 1953. J. Biol. Chem. **200:** 493.
12. King, T. P. & L. C. Craig. 1958. J. Am. Chem. Soc. **80:** 3366.
13. Abel, J. J. 1926. Proc. Natl. Acad. Sci. U. S. **12:** 132.
14. Craig, L. C. 1956. *In* Internal Secretions of the Pancreas, Ciba Foundation Colloquia. **9:** 107. Little, Brown & Co. Boston, Mass.
15. Yphantis, D. A. & D. F. Waugh. 1957. Biochim. et Biophys. Acta. **26:** 218.
16. Fredericq, E. 1956. *In* Internal Secretions of the Pancreas, Ciba Foundation Colloquia. **9:** 89. Little, Brown & Co. Boston, Mass.
17. Waugh, D. F. 1956. *In* Internal Secretions of the Pancreas, Ciba Foundation Colloquia. **9:** 122. Little, Brown & Co. Boston, Mass.
18. Laskowski, M., S. J. Leach & H. A. Scheraga. 1960. J. Am. Chem. Soc. **82:** 571.

RAPID DETERMINATION OF MOLECULAR WEIGHTS OF PEPTIDES AND PROTEINS*

David A. Yphantis
The Rockefeller Institute, New York, N. Y.

The determination of molecular weight by equilibrium ultracentrifugation has a number of advantages: first, the thermodynamic basis of the technique is quite secure; second, it can handle a very broad range of molecular weights; and third, it requires only small amounts of material. The major disadvantage of the technique has been the long time required to reach complete equilibrium: for example, it has been common to run the ultracentrifuge for more than one week to attain equilibrium with proteins.

Recently Van Holde and Baldwin[1] proposed the use of short columns of solution, that is, small distances between meniscus and base, to reduce the time required. In experiments with ribonuclease in 3 mm. high columns of solution, they found complete equilibrium in 14 hours and calculated excellent values for the molecular weight. They also studied sucrose in both 3 mm. and 1 mm. columns. The values obtained for the molecular weight of sucrose in the 3 mm. columns were in very good agreement with theory, but they observed a small systematic error in the 1 mm. columns.

Under practical conditions, the time required to attain equilibrium is inversely proportional to the square of the column height. Thus one could decrease the equilibrium time considerably by using even shorter columns. It was felt worthwhile, therefore, to repeat the experiments of Van Holde and Baldwin with shorter columns.

Short Column Method

The differential equation for sedimentation equilibrium of a two-component system, taking into account concentration but not pressure dependence, is

$$\frac{1}{rc}\frac{dc}{dr} = \frac{\omega^2 M(1 - \bar{V}\rho)}{RT\left(1 + c\dfrac{d \ln y}{dc}\right)} \tag{1}$$

Here r is the distance from the center of rotation, c is the concentration, ω is the angular velocity, M is the molecular weight, $\bar{V}$ is the partial specific volume of the solute, y is the activity coefficient of the solute on the c scale, ρ is the density of the solution, R is the gas constant, and T is the absolute temperature. With peptides and proteins one can usually choose conditions to minimize concentration dependence and make the term $c\, d \ln y/dc$ negligible.

If the position in the column where the concentration was equal to the original concentration c_o were known, the concentration gradient $(dc/dr)_o$ at that point could be determined, and thus the right-hand side of the equation could be evaluated. Van Holde and Baldwin[1] showed that at the point

$$r' = \sqrt{(a^2 + b^2)/2},$$

* The work described in this paper was supported in part by Grant A-2493 from the National Institute of Arthritis and Metabolic Diseases, Public Health Service, Bethesda, Md.

which for a 1-mm. column is within 0.002 mm. of the midpoint, the concentration is given by

$$c(r') = c_o H/\sinh H = c_o/(1 + H^2/6 + \cdots) \qquad (2)$$
$$H = \omega^2 M(1 - \bar{V}\rho)(b^2 - a^2)/4RT$$

Here a and b are the distances from the center of rotation to the meniscus and base of the solution. The concentration at the mid-point of a short column is within 1 per cent of the initial concentration if $H < 0.25$. For ribonuclease in an 0.7 mm. column, this corresponds to speeds under 24,000 rpm and, in a 3-mm. column, it corresponds to speeds under 11,600 rpm. Thus for sufficiently low speeds in short columns one can simply determine the initial concentration and the concentration gradientat the midpoint of the column and evaluate a molecular weight from EQUATION 1.

Materials

Sucrose and buffer salts were of the analytical reagent grade. Crystalline bovine ribonuclease was from several samples of one lot.* Two samples of bovine ribonuclease A were used. These samples were isolated by countercurrent distribution[2] from crystalline bovine ribonuclease,† and were more than 99 per cent pure, as determined by column chromatography, countercurrent distribution, amino acid analyses, and nitrogen analysis. Crystalline bovine β-lactoglobulin‡ was also used. The proteins were dissolved in, and exhaustingly dialyzed against, a *p*H 7.75 buffer containing 0.1 *M* NaCl, 0.035 *M* K_2HPO_4 and 0.004 *M* KH_2PO_4. The lower concentrations in an experiment were prepared by dilution by weight from the most concentrated solution after dialysis. The absolute concentrations indicated for the proteins are approximate.

Equipment

The ultracentrifuge used§ was equipped with a phase plate and with an automatic temperature control unit calibrated against a National Bureau of Standards-certified thermometer. Both schlieren and Rayleigh interference optics were used. The optics were modified by placing the limiting aperture over the lower collimating lens on a mount designed to allow both lateral and angular adjustments. The light source was carefully aligned to be in the focal plane of the first collimating lens and accurately on axis by a reflection technique similar to that of Trautman.[3] The alignment procedure of Richards and Schachman[4] was followed with the following modifications. First, with only the standard schlieren limiting aperture in place on the drive fork, the cylindrical lens was tilted for maximum sharpness of the meniscus image of a water-filled cell spun at about 15,000 rpm. Then, after setting the interference position for the light source, the interference limiting aperture was aligned so that it would be centered for all radial positions in the cell: a standard double channel cell with the center rib removed was filled with a 1 per cent solution

* Armour Lot 381-059.
† Armour Lots 381-059 and 381-062.
‡ Pentex Lot 4813.
§ Spinco model E, Spinco Div., Beckman Instruments Corp., Palo Alto, Calif.

of ribonuclease and 0.1 ml. of Fluorocarbon FC-43* to allow observation of the base region. The cell was run at a speed of about 20,000 rpm to develop concentration gradients at the meniscus and base, and was then slowed to about 5000 rpm. The distribution was allowed to diffuse a few minutes at this latter speed. Lateral and angular adjustments of the limiting aperture were made so that the Rayleigh interference fringes were straight throughout the cell. This procedure aligns the limiting aperture so that it is centered throughout the cell and eliminates the possibility of alignment at one region only, which is possible if the limiting aperture is aligned with a boundary near the center of the cell.

All schlieren observations were made at a constant phase-plate angle of 80°. Exposures were taken on spectrographic plates, emulsion type IIG.† The plates were measured with a two dimensional comparator.‡ At least five fringes, three light and two dark, were read for every determination of fringe position on the interference records. After subtraction of the appropriate number of half-fringe spacings, the readings were averaged and the standard deviation of a measurement was calculated; on good plates this was usually less than 5 μ. The thickness of every centerpiece was measured and appropriate corrections were made for any variations. The double-channel synthetic boundary cells were either the Kegeles type[5] or the capillary type supplied by the manufacturer. Synthetic boundary cell runs were performed only on solutions of concentration ≥ 0.5 per cent to avoid boundary instabilities at low concentrations.

Calculations

With the standard schlieren optics, the initial concentration was determined in terms of arbitrary refractive index units by integration of the synthetic boundary trace.[6,7] The value of the refractive index gradient at the mid-point of the cell was given directly by the difference between the solution and solvent traces. For reliable measurements, the solution and solvent patterns should be distinctly separated, or gross errors may result. With the interference optics, the concentration was given directly by the number of fringes across the synthetic boundary,[4] and the concentration gradient at the mid-point of the short columns was determined in either of two ways. The simplest way was to use the fringe difference between two points equidistant from the middle of the column. Alternatively, a quadratic or cubic equation with the mid-point as the origin was fitted by least squares to the fringe positions at five to nine points symmetrically disposed about the mid-point of the column; the coefficient of the linear term then gave the value of the concentration gradient at the mid-point. Values of the radius to a point were corrected for rotor expansion.[8] Only in the case of sucrose were there any values of $c\ d\ \ln y/dc$ available.[9] The usual practice of assuming activity coefficient terms of unity was followed with the proteins, so that the apparent molecular weight was calculated from

$$M = \frac{1}{\bar{r}c_o}\left(\frac{dc}{dr}\right)_{r=\bar{r}} \frac{RT}{\omega^2(1 - \bar{V}\rho)} \quad \ldots \quad (3)$$

* Perfluorotributyl amine, a volatile, dense, and inert liquid supplied by the Minnesota Mining and Manufacturing Co., St. Paul, Minn.

† Kodak.

‡ Model No. M2001RS, manufactured by the Gaertner Scientific Corp., Chicago, Ill.

where $\bar{r} = (a + b)/2$. The following values were used for the partial specific volumes: sucrose,[9] $\bar{V} = 0.618$; ribonuclease,[1,10] $\bar{V} = 0.695$; and β-lactoglobulin,[11] $\bar{V} = 0.7514$. The densities of the sucrose solutions were interpolated from the data of Morris and Gosting,[9] while the density of the buffer was esti-

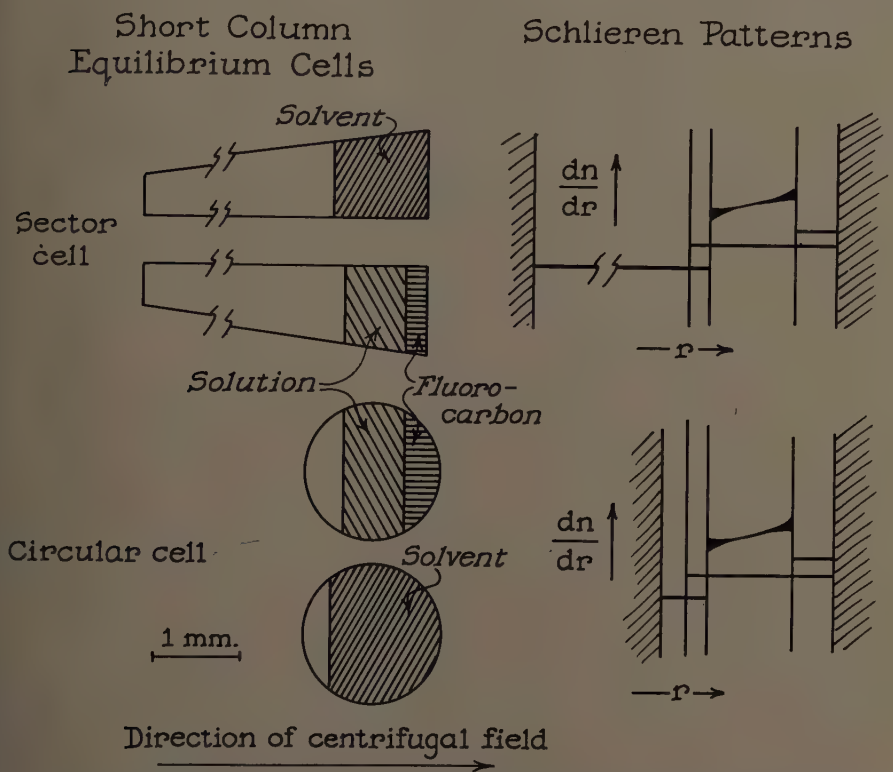


FIGURE 1. Comparison of short column equilibrium ultracentrifugation in sector and circular channels, illustrating the loading of the channels and the resultant schlieren diagrams. The one-mm. scale does not apply to the lateral dimensions of the sector cell.

mated from the *International Critical Tables*. For the protein solutions, the densities were calculated by adding the density increment due to the protein

$$\Delta\rho_p = C_o(1 - \bar{V}\rho_b) \qquad (4)$$

to the buffer density. In EQUATION 4, C_o is the protein concentration of the initial solution in gm. ml.$^{-1}$, and ρ_b is the buffer density.

Sector Channel Cells

The experiments of Van Holde and Baldwin[1] were repeated using the standard double-channel cells and schlieren optics. One sector channel was loaded with 0.025 ml. of FC 43 and varying amounts of solution; the other channel was loaded with solvent to a slightly higher level (FIGURE 1). The solvent

volume is not critical unless there is significant redistribution of solvent components. The cells were run and photographed periodically until after equilibrium had been reached, as evidenced by the lack of further changes. The patterns observed were quite indistinct near the ends of the columns, presumably because of diffraction. Reliable measurements could be made on the schlieren patterns, however, at points further than 0.5 to 0.6 mm. on the plate from the meniscus and base images if sufficiently high phase-plate angles (sufficiently low vertical magnification) were used. Thus the refractive index gradient apparently could be determined for points further than about 0.25 mm. from either of the ends of the cell and the refractive index gradient determined at the mid-point of columns longer than 0.5 to 0.6 mm. The use of approximately 0.7-mm. columns and a phase-plate angle of 80° appears to give adequate definition to the schlieren trace at the mid-point of the column with good sensitivity. Even higher phase-plate angles and somewhat shorter columns could be used at the sacrifice of sensitivity, or longer columns and lower phase-plate angles at the sacrifice of equilibrium time.

Under these conditions runs were made on a sample (No. I) of crystalline bovine ribonuclease and on a sample of bovine ribonuclease A at concentrations of approximately 0.5 and 1.0 per cent and at speeds of 12,590 and 15,220 rpm. Equilibrium was established in about 40 min. The molecular weights calculated ranged from 13,510 to 13,940 with standard deviations of about 2 per cent on each determination. These values are to be compared with the formula weight of 13,683.[12]

There are two primary disadvantages to the use of such small columns: (1) the results are less precise than can be obtained with the longer columns,[1,4] although still adequate for many applications; and (2) there is considerably less sensitivity to the heterogeneity of a sample than with the longer columns. Thus, by using short columns, one trades precision and resolving power for time.

In addition to decreasing the running time, the use of short columns has two more advantages: (1) Higher speeds may be used, while still maintaining the mid-point concentration close to the original concentration. This allows the observation of more dilute solutions. For example, a 0.75-mm. column of solution may be operated at twice the speed possible with a 3-mm. column; thus, in principle at least, it should be possible to study solutions 4 times less concentrated. Another advantage of short columns is their compactness. This allows the design of multichannel cells for the simultaneous observation of a number of solutions.

Circular Channel Cells

It is easier to drill a round hole than to machine a sector or even a rectangular hole. Since the form of the equilibrium distribution for a monodisperse ideal solute is independent of the shape of the cell, the possibility of using circular channels was investigated. If a channel of circular cross section is filled with solution so that the meniscus and base of the solution column are the same distance (h) from the diameter (l) of the channel (FIGURE 2*a*), then the concentration at the mid-point of the channel (and of the column) may be shown to

approximate the initial concentration. The form of the concentration distribution for an ideal monodisperse solute at sedimentation equilibrium is given by

$$c = A' \exp \langle \omega^2 M(1 - \bar{V}\rho)r^2/2RT \rangle \quad (5)$$

For simplicity in short columns this may be approximated by

$$c = A' \exp \langle \omega^2 M(1 - \bar{V}\rho)\bar{r}^2/2RT \rangle \exp \langle \omega^2 M(1 - \bar{V}\rho)\bar{r}x/RT \rangle = Ae^{kx} \quad (6)$$

where A' and A are constants reflecting, in part, the effects of the geometry of the column, $\bar{r} = (a + b)/2$ is the radius to the mid-point of the column, $x = \bar{r} - r$ is the distance from the mid-point of the column, and $k = \omega^2 M(1 - \bar{V}\rho)\bar{r}/RT$. This approximation involves the neglect of a factor $\exp \langle \omega^2 M(1 - \bar{V}\rho)x^2/RT \rangle$, which for $x \leq 0.05$ cm. and under the most extreme useful operat-

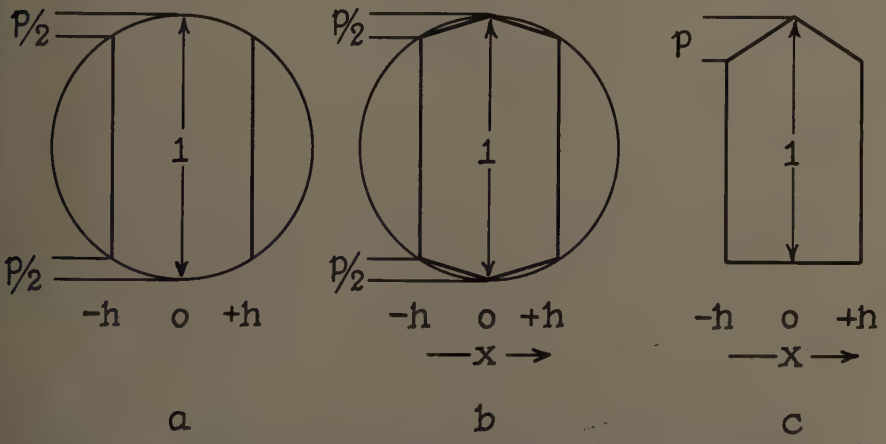


FIGURE 2. (*a*) Solution volume in circular channel. (*b*) Chord approximation. (*c*) Equivalent cross section. See text for details.

ing conditions differs by less than 0.0025 from unity. The concentration at the mid-point of the column where $x = 0$ is given by the value of A. This may be determined by equating the amount of material present in the solution column at equilibrium with that present initially. It is convenient to approximate the circular walls by straight ones, as indicated by the chords in FIGURE 2*b*, so that the solution column is represented by two symmetrical trapezoids and is equivalent in cross section to the shape shown in FIGURE 2*c*. One could improve the approximation to the circular walls by using more chords, but this appears unnecessary.

The amount of material initially present in the channel is given by

$$q = c_o z \left(\int_{-h}^{h} l \, dx + \int_{-h}^{o} \frac{px}{h} dx - \int_{o}^{h} \frac{px}{h} dx \right) = c_o z(2hl - hp) \quad (7)$$

where z is the thickness of the channel and c_o is the original concentration. At equilibrium, the amount of material in the channel is the same and is given by

$$q = Az\left(\int_{-h}^{h} le^{kx}\,dx + \int_{-h}^{o} \frac{pe^{kx}x\,dx}{h} - \int_{o}^{h} \frac{pe^{kx}x\,dx}{h}\right)$$
$$= Az\left(\frac{2l}{R}\sinh kh + \frac{2p}{hk^2}\langle\cosh kh - 1 - kh\sinh kh\rangle\right) \quad (8)$$

By eliminating q from Equations (7) and (8), expanding functions in power series up to the fourth power, solving for A and, finally, omitting terms in $(kh)^4$ and higher powers, it can be shown that

$$A = c(\bar{r})/c_o = 1 - \frac{k^2h^2}{6}\left(1 - \frac{p}{4(l - p/2)}\right) + \cdots \quad (9)$$

If the circular channel is filled so that the height of the solution columns $2h$ is equal to the radius of the channel $l/2$, then

$$c(\bar{r})/c_o = 1 - 0.964\,k^2h^2/6 + \cdots \quad (10)$$

In this case, and even more so for shorter columns, the concentration at the mid-point of the circular column is, for almost all purposes, identical with that at the mid-point of a rectangular or sector-shaped column, where

$$c(\bar{r})/c_o = 1 - k^2h^2/6 + \cdots \quad (11)$$

Thus, in so far as the mid-point concentration is concerned, a circular channel can be made equivalent to a sector channel. The geometry assumed for the above derivation may be realized readily by placing appropriate amounts of solution and base fluid in the channel. It can be shown, both theoretically and experimentally, that the conditions outlined above are more stringent than are actually required. A variation of ±50 per cent in the amount of base fluid has no detectable effect (<2 per cent) on the calculated value of the molecular weight even for the highest useful speeds.

The arrangement for a circular channel cell and the resulting schlieren patterns are illustrated in FIGURE 1*b*. One circular cross-section channel is filled with the approximate amounts of base fluid and solution required by the above considerations. Another circular channel, at the same distance from the center of rotation, is filled with somewhat more solvent than the total of solution and base fluid in the first channel (should buffer redistribution be significant, base fluid can be added also to this reference channel to secure equivalent conditions in the two channels). These two channels are then equivalent to a double sector cell for equilibrium runs calculated as described previously.

Multichannel Cells

Such pairs of channels can be grouped together into a single ultracentrifuge cell to give multichannel cells. Cells with eight circular cross-section observation channels (FIGURE 3) were designed and machined from duraluminum-filled epoxy centerpiece blanks.* Four of these channels serve for solution and the other four for solvent, thus providing a reference base line for each solution. These eight observation channels are one sixteenth of one inch diameter holes extending all the way through the centerpiece. The principle of a side pocket

* Kindly supplied by the Spinco Division, Beckman Instrument Corp.

reservoir embodied in the synthetic boundary cell of Kegeles[5] was used as a method of filling the cells. Eight blind holes, three thirty-seconds of one inch in diameter, served as filling reservoirs. They were drilled at an angle of about 2° to the observation holes in the direction of the centrifugal field so that they would tend to empty completely. Small grooves were cut on the top face of the centerpiece to connect each observation channel with its appropriate filling reservoir. On acceleration the contents of the filling reservoirs drained through the grooves and into the observation channels where the equilibrium distribution was set up. The top and bottom faces of the centerpieces were lapped with fine emery paper to remove tool marks. This should be done very carefully to avoid irregularities and unevenness.

Cells with six observation channels also were made. The channels in these cells were displaced one sixteenth of one inch so that one of these cells could be

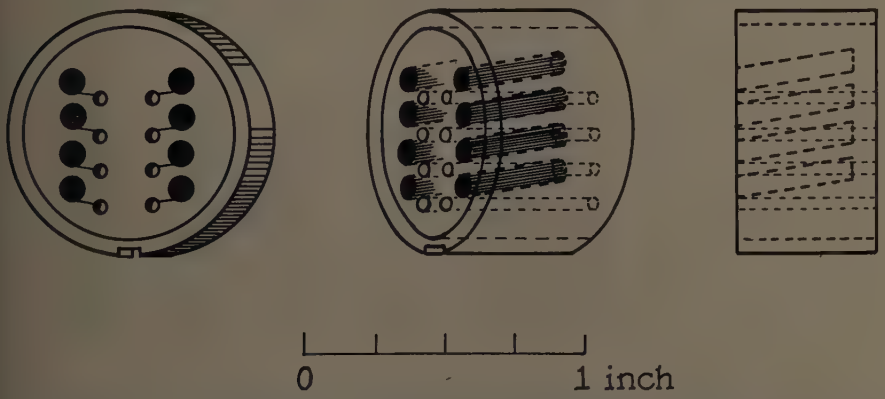


FIGURE 3. Eight-channel cell. Direction of centrifugal field is toward the foot of the page. The light holes are observation channels and the dark holes are filling reservoirs. On acceleration the contents of a filling reservoir drain through a groove into the corresponding observation channel.

used as a counterbalance to an eight-place cell and thus allow simultaneous observation of seven solutions in the standard two-place rotors. Using wedge windows with the schlieren optics or wedge cells with the interference optics as many as 42 solutions in the 6-place rotor can be observed simultaneously.

The filling of the cells requires some care. All solutions were handled with 0.05-ml. microsyringes* with a piece of drawn-out No. 20 Kel-F† tubing worked into the delivery hole of the syringe. This Kel-F tubing was long enough so that the solutions never contacted any parts of the syringe itself. The same syringe can be used repeatedly without contamination simply by changing the Kel-F tubing. As a precaution, the tubing was rinsed with solution before use. The filling reservoirs on one side of the centerpiece were first filled with 0.005 ml. of FC-43. Then the filling reservoirs on the opposite side were filled with 0.021 ml. of solvent. Finally the solutions, usually 0.015 ml., were loaded into

* Hamilton Co., Whittier, Calif.
† Resistoflex Corp., Roseland, N. J.

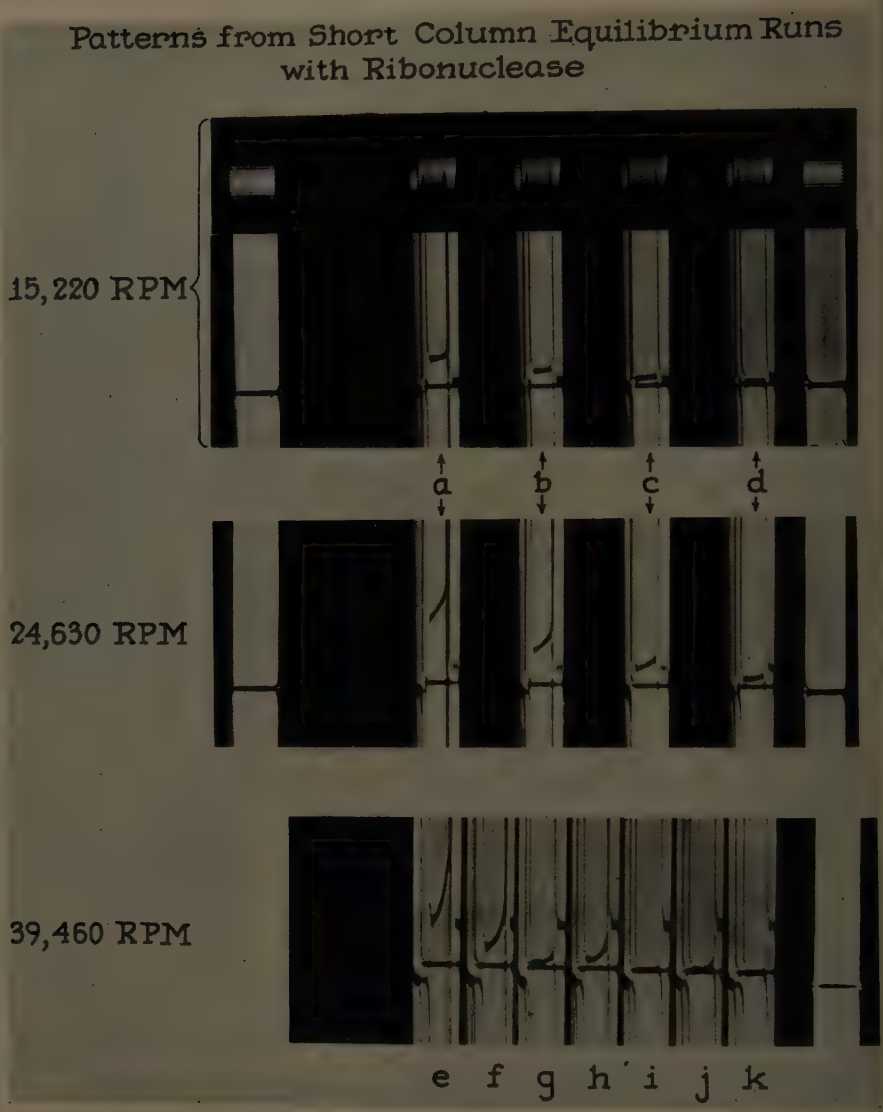


FIGURE 4. Typical equilibrium patterns with ribonuclease in multichannel cells. The top pattern was taken with the Rayleigh interference optics, the others with the schlieren optics at an 80° phase-plate angle. Concentrations of ribonuclease in channels: (*a*) 1 per cent, (*b*) ½ per cent, (*c*) ¼ per cent, (*d*) ⅛ per cent, (*e*) ½ per cent, (*f*) ⅕ per cent, (*g*) 1/20 per cent, (*h*) 1/10 per cent, (*i*) 1/100 per cent, (*j*) 1/40 per cent, and (*k*) 0 per cent. The bottom pattern was obtained using a six-channel cell and an eight-channel cell simultaneously, while the other patterns were obtained using an eight-channel cell alone.

the first set of filling reservoirs and the cell was assembled in the normal manner. Rapidity of filling and assembly is essential to minimize the effects of evaporation with such small volumes; the filling procedure above minimizes the exposure of the solutions to the atmosphere.

Typical equilibrium patterns with these cells are shown in FIGURE 4 for several concentrations of ribonuclease at various speeds. A comparison of the schlieren and interference patterns at 15,220 rpm shows that the interference and schlieren optical systems are complementary. The interference optics appear useful with the lower refractive index gradients where the schlieren traces of the solution and solvent overlap, while the schlieren optics can be used with the

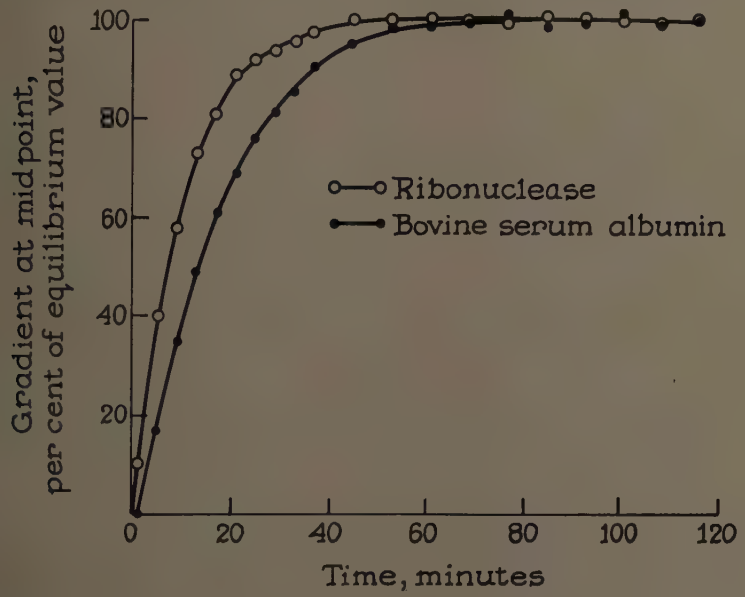


FIGURE 5. Approach to equilibrium for ribonuclease and bovine serum albumin at 12,590 rpm.

larger gradients where the Rayleigh fringes are not resolved. The bottom record shows the arrangement of the images, using an eight-channel and a six-channel cell simultaneously.

Equilibrium Time

The time required to attain equilibrium in the circular channel cell is essentially the same as in the sector cell. FIGURE 5 presents the refractive index gradient at the mid-point of 0.8-mm. columns as a function of time for ribonuclease and bovine serum albumin at a speed of 12,590 rpm. The equilibrium times are seen to be about 45 and 70 min. respectively. With sucrose, equilibrium was obtained in about 15 min. Although the initial values of the refractive index gradient in the sector and circular channel cells at the mid-point differed somewhat, possibly because of convective transport in the circular cells,

essentially the same values were obtained for both types of cells at the later times. Thus the time required to approach equilibrium in the circular channels can be estimated from the equations of Van Holde and Baldwin.[1]

Results with Sucrose

In spite of its low molecular weight, sucrose appears to be an excellent test substance for equilibrium ultracentrifugation since there are values available for its activity coefficient over a wide concentration range. TABLE 1 gives the

TABLE 1

MOLECULAR WEIGHT OF SUCROSE DETERMINED IN THE MULTICHANNEL SHORT COLUMN EQUILIBRIUM CELL, USING SCHLIEREN OPTICS

Sucrose concentration gm. 100 ml.	M_{obs}^*/M at 42,040 rpm	M_{obs}^*/M at 50,740 rpm	M_{obs}^*/M at 59,780 rpm
1.00	—	—	1.011 ± 0.013
1.68	0.974 ± 0.029	0.999 ± 0.012	1.021 ± 0.013
2.00	1.032 ± 0.029	—	1.009 ± 0.016
2.52	1.000 ± 0.018	0.996 ± 0.017	0.998 ± 0.013
3.00	0.994 ± 0.009	—	0.995 ± 0.013
3.36	1.018 ± 0.018	1.005 ± 0.013	1.013 ± 0.016
4.20	1.035 ± 0.015	0.964 ± 0.013	0.991 ± 0.014

* M_{obs} is the experimentally determined molecular weight, corrected for nonideality, as calculated from EQUATION 1. M is the known molecular weight of sucrose, 342.3.

TABLE 2

MOLECULAR WEIGHT OF SUCROSE DETERMINED IN THE MULTICHANNEL SHORT COLUMN EQUILIBRIUM CELL, USING RAYLEIGH INTERFERENCE OPTICS

Sucrose concentration gm./100 ml.	M_{obs}^*/M at 29,500 rpm	M_{obs}^*/M at 42,040 rpm	M_{obs}^*/M at 59,780 rpm
0.50	0.98 ± 0.11	1.043 ± 0.050	0.990 ± 0.050
1.00	0.96 ± 0.09	1.024 ± 0.043	1.005 ± 0.035
2.00	0.976 ± 0.047	0.995 ± 0.015	—
3.00	0.976 ± 0.026	0.972 ± 0.022	—

* M_{obs} is the experimentally determined molecular weight, corrected for nonideality, as calculated from EQUATION 1. M is the known molecular weight of sucrose, 342.3.

molecular weights obtained with sucrose in the multichannel cells, using the standard schlieren optics. The ratio of the molecular weight calculated by means of EQUATION 1 to the known molecular weight of 342.3 is presented for various speeds and concentrations along with the standard deviations from the measurements. Again, the precision of the results is not as high as can be obtained with the longer columns, but the results appear reliable to about 3 per cent.

TABLE 2 presents values for the molecular weight of sucrose, using the interference optics. At the higher speeds the usual quartz windows distort significantly and the interference pattern deteriorates badly. Therefore the quartz windows were replaced with the recently available sapphire windows. Al-

though these sapphire windows are, at present, optically inferior at low speeds, they distort much less and give usable patterns at the higher speeds. The results are seen to be dependable to within about 3 per cent provided there is a reasonable gradient in the cell. With the lower gradients, the precision, of course, falls off; for example, in the run with ½ per cent sucrose at 29,500 rpm the total fringe shift across 0.45 mm. of the cell was only 58 ± 6 μ.

Ribonuclease

A common test protein for the demonstration of methods of molecular weight determination has been crystalline bovine ribonuclease. The usual material is an Armour preparation, sometimes of unspecified lot number[7,13] and other

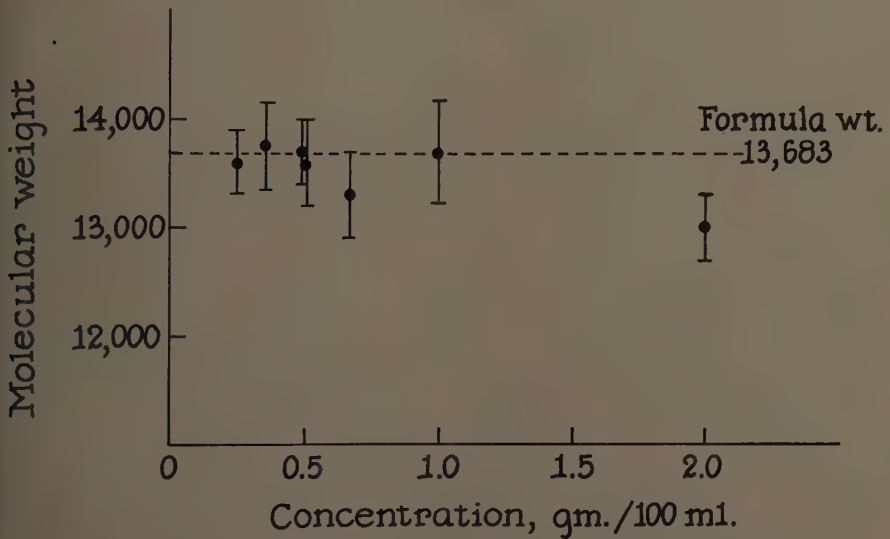


FIGURE 6. Molecular weight of ribonuclease as a function of concentration.

times of Lot No. 381-059.[1,3,14,15] Several samples of this latter lot, purchased at different times, were examined. These samples were stored, well sealed at −10° C., and were warmed to room temperature before opening. No other precautions were taken to exclude moisture.

FIGURE 6 presents the molecular weights determined for a freshly opened sample (No. II) of Lot No. 381-059 as a function of concentration. These results were obtained with the schlieren optics at speeds of 15,220 and 21,740 rpm. It may be seen that the results are in good agreement with the formula weight. The precision of the results is about 3 per cent. Essentially the same values were obtained with a sample of ribonuclease A. Sample II was used for various purposes, and when the molecular weight was redetermined five months later, a value of 14,700 ± 250 was found at a concentration of one per cent and at a speed of 15,220 rpm. Two other samples (III and IV) of Lot No. 381-059 gave initial values of 14,250 ± 250. Again, higher values, up to 14,850 ± 300, were found after the samples had been used over a period of months. These high

values were confirmed on one sample by equilibrium ultracentrifugation in a 3-mm. column. Molecular weight heterogeneity of ribonuclease samples has been reported previously.[4,16]

A single experiment was performed on sample II when it was freshly opened, using the interference optics. The values found for the molecular weight were 13,850 ± 270 and 13,770 ± 150 for 0.5 per cent and 1 per cent solutions at 15,220 rpm and 13,400 ± 150 for an 0.5 per cent solution at 21,740 rpm. Several runs were made later with various samples at lower concentrations, from 0.1 per cent to 0.01 per cent. The molecular weights were found to decrease as the concentration was lowered, with the 0.01 per cent solution giving an apparent value for the molecular weight of 5200 ± 1500. In view of the known covalent structure of ribonuclease, these results were interpreted in terms of adsorption onto the walls of the cells with a lowering of the effective initial concentration. In retrospect, such adsorption is to be expected with the high surface-to-volume ratio in these cells (~1.6 cm.^{2}/0.015 cm.^{3} sample volume) and with the rough surfaces of the machined epoxy resin. Two methods for minimizing the effects of such adsorption are obvious: redesign of the cells to decrease the surface-to-volume ratio and modification of the surfaces in contact with the protein solutions. Both of these are being investigated. Eight-channel cells machined from Kel-F are under test, and appear promising for use at the lower speeds and also for use with organic solvents and extremes of *p*H.

Adsorption lowers the effective concentration in the cell but should not influence the form of the equilibrium distribution, assuming uniform adsorption of the various components of a solution. Techniques of calculation not involving values of the original concentrations should still be applicable in the presence of adsorption. Such a calculation method has been given by Van Holde and Baldwin,[1] who showed that the slope of the plot of $(dn/dr)/r$ *versus* n yields the Z average* molecular weight, where n is the refractive index of the solution. In such short cells as these, the variation in r is negligible compared to the experimental errors, so a plot of (dn/dr) *versus* n should be equivalent. Attempts were made to use the schlieren optics for such determinations with 0.7- to 0.8-mm. columns, but the results were disappointing. At best, the standard deviations were large (~30 per cent), and the molecular weights were reliable to only 20 per cent. Using the interference optics, reliable measurements apparently can be made much closer to the ends of the cell (to within 0.15 mm. of a meniscus image on the plates)[16] as compared with 0.5 to 0.6 mm. with the schlieren optics. Thus with the interference optics it should be possible to obtain values over more of the solution column. Accordingly, some runs were made with the interference optics at speeds high enough to give a distinct curvature to the Rayleigh fringes.

The following values were found for M_z the Z average molecular weight of the second sample II of the ribonuclease: 14,700 ± 1700 at a speed of 25,980 rpm with a concentration of 0.5 per cent; 14,800 ± 800 at 39,460 rpm with 0.1 per cent; and 15,000 ± 1500 at 39,460 rpm with 0.05 per cent. A least-squares

* Strictly speaking, this is their $\bar{M}_{2/1}$ that reduces to the Z average if the refractive increments and partial specific volumes of the components are identical.

treatment of the data from 5 positions in the 0.7-mm. columns was used to determine these values and their indicated standard deviations. The increased sensitivity at the higher speeds is due to the greater curvature of the fringes; the value of M_z is essentially proportional to the second derivative of the concentration. For such short columns as these, the values obtained for $\overline{M}_z$ can be considered only as estimates, with an accuracy of about 10 per cent. More precise values appear to require longer columns and longer equilibrium times.

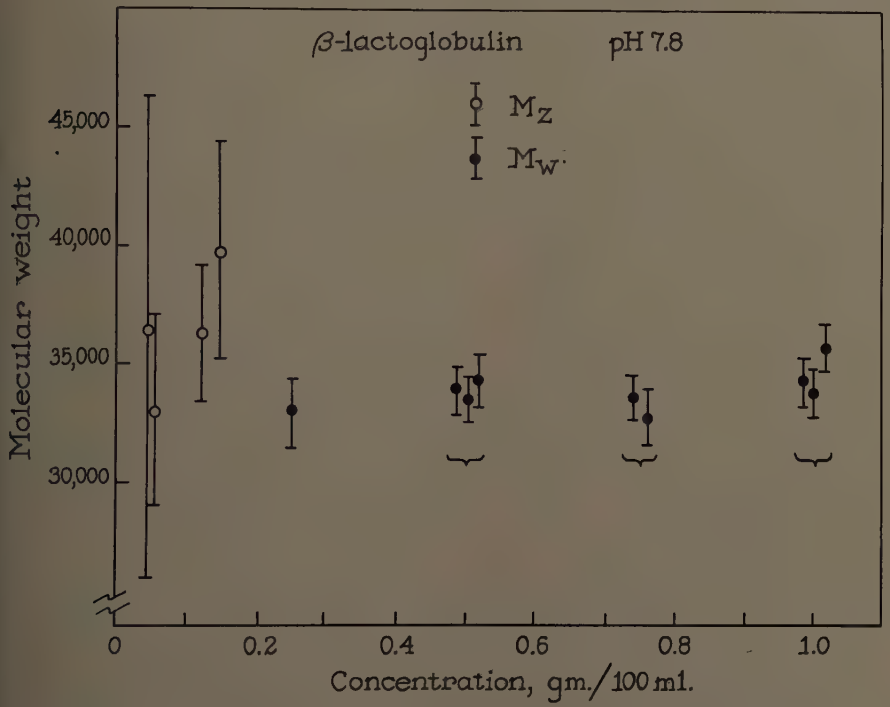


FIGURE 7. Molecular weight of β-lactoglobulin as a function of concentration.

β-Lactoglobulin

FIGURE 7 presents values found for the molecular weight of a commercial sample of β-lactoglobulin. The open circles give the values obtained from the refractive index gradients at the mid-point, using the schlieren optics and a speed of 12,590 rpm. Runs at concentrations of 0.15 per cent and less gave progressively lower values. Values of the Z average molecular weight determined at 25,980 rpm are also included, along with a single point of poor precision, obtained with a 0.05 per cent solution at 35,600 rpm. Although the curvature of the fringes was large in this case, the fringes were spaced too closely for resolution near the base of the solution column. Since the Z average molecular weights remained in the neighborhood of 35,000, these results with β-lactoglobulin were again interpreted in terms of adsorption.

Discussion

The molecular weights determined using the gradient at the mid-point of the columns are essentially the weight average molecular weights. At speeds so low that all the components are at the original concentration at the mid-point of the column, the molecular weight average is the M_1 of Van Holde and Baldwin.[1] If all the components have the same refractive increments and the same partial specific volumes, this is exactly the weight-average molecular weight. As the equilibrium speed is increased, the concentrations of the heaviest components decrease at the mid-point, thus lowering the apparent molecular weight. The presence of a small percentage of material with a molecular weight much higher than the main components can be discerned easily by a steep rise in the schlieren pattern at the base of the column or by a thickening of the interface between solution and base fluid. The ratio of the Z average molecular weight M_z to the weight average molecular weight $\bar{M}_w$ is an index of the dispersity of the sample. The standard deviation of the molecular weight distribution curve is given by

$$\sigma_M = \bar{M}_w \sqrt{\bar{M}_z/\bar{M}_w - 1}. \tag{12}$$

Unfortunately, these short column runs do not give precise values for $\bar{M}_z$. At best the results appear subject to about a 10 per cent error. Thus σ_M can be calculated only to about 8 per cent. For more precise determinations one should use longer columns.

The present low concentration limit of this short column technique appears to be about 0.2 per cent with the epoxy cells. At lower concentrations, the weight-average molecular weights determined for ribonuclease and β-lactoglobulin appear to drop. This has been interpreted as a result of adsorption of the proteins onto the cell walls. In preliminary experiments, Kel-F cells appear promising for lower concentrations, but only for speeds under 30,000 rpm, since they are much weaker. The amount of material required for a molecular weight determination is small, and most of it is required for the synthetic boundary run. With care, a weight average molecular weight determination requires about 0.6 mg. of material. The concentration dependence of the molecular weight can be determined in the range of 0.2 to 1 per cent with less than 2 mg. If a 10 per cent error is acceptable, the Z average molecular weight can be determined with as little as 8 μg. of material, under favorable conditions.

The range of molecular weights that can be determined is set by the speed limits and optical sensitivity of the ultracentrifuge. With one per cent solutions, the molecular weight of amino acids could probably be estimated to about 10 per cent. The largest molecule examined to date has been a sample of serum albumin dimer. It should be possible to use this technique with much larger molecules. With a recent heavy rotor it is possible to run the ultracentrifuge without disturbing precession or vibration at speeds below 2000 rpm. This should allow the determination of the molecular weights of the smaller viruses with an equilibrium time of about 8 hours.

This technique has been in routine use for the past few months. The materials examined include oligopeptides, antibiotic polypeptides, various samples

of proteins such as insulin, ribonuclease, pepsin and albumins, and soluble ribonucleic acids. The technique appears most useful as a survey tool for the rapid examination of a number of samples or different conditions. For many problems the precision of 3 per cent is sufficient; for example, a partial specific volume from the amino acid composition often is assumed or calculated. With these multichannel cells the various concentrations are automatically run at the same temperature and essentially the same pressure. The short time required for equilibrium allows several runs to be made on the same day at different temperatures. Thus the method seems promising for the determination of the thermodynamic parameters of association-dissociation reactions.

Acknowledgments

I express my appreciation to Owen Griffith for his technical assistance, to H. K. Schachman for a prepublication copy of a manuscript,[4] and to L. G. Longsworth for many helpful discussions. I am also grateful to T. P. King and David Eaker of The Rockefeller Institute, who generously supplied the samples of bovine ribonuclease A mentioned under *Materials*.

References

1. Van Holde, K. E. & R. L. Baldwin. 1958. J. Phys. Chem. **62:** 734.
2. King, T. P. & L. C. Craig. 1958. J. Am. Chem. Soc. **80:** 3366.
3. Trautman, R. 1958. Biochim. et Biophys. Acta. **28:** 417.
4. Richards, E. G. & H. K. Schachman. 1959. J. Phys. Chem. **63:** 1578.
5. Kegeles, G. 1952. J. Am. Chem. Soc. **74:** 5532.
6. Klainer, S. M. & G. Kegeles. 1955. J. Phys. Chem. **59:** 952.
7. Ginsburg, A., P. Appel & H. K. Schachman. 1956. Arch. Biochem. Biophys. **65:** 545.
8. Schachman, H. K. 1959. Ultracentrifugation in Biochemistry. Academic Press. New York, N. Y.
9. Gosting, L. J. & M. S. Morris. 1949. J. Am. Chem. Soc. **71:** 1998.
10. Harrington, W. F. & J. A. Schellman. 1956. Compt. rend. trav. lab. Carlsberg Ser. chim. **30:** 21.
11. Pedersen, K. O. 1936. Biochem. J. **30:** 961.
12. Hirs, C. H. W., W. H. Stein & S. Moore. 1956. J. Biol. Chem. **221:** 151.
13. Erlander, S. R. & J. F. Foster. 1959. J. Polymer Sci. **37:** 103.
14. Klainer, S. M. & G. Kegeles. 1956. Arch. Biochem. Biophys. **63:** 247.
15. Trautman, R. & C. F. Crapton. 1959. J. Am. Chem. Soc. **81:** 4036.
16. Yphantis, D. A. 1959. J. Phys. Chem. **63:** 1742.

Part II. Degradation and Structure

PHENYLTHIOHYDANTOINS IN PROTEIN ANALYSIS

Pehr Edman
St. Vincent's School of Medical Research, Melbourne, Australia

Some years ago I described a procedure (P. Edman, 1949, 1950, 1953) for the stepwise degradation of simple peptides using phenylthiohydantoins (PTH). In principle, this degradation reaction should work for amino acid sequences of any length and variety provided the amino acids are linked by normal peptide bonds. However, this extension has not come easily. In view of its importance for development in the field of protein structure, I have devoted some efforts to bringing it about. My presentation will be concerned chiefly with this problem, but I shall also draw attention to other applications of the PTH technique in protein analysis.

Several problems had to be solved before it became feasible to attack large structures, and they will be dealt with in turn. First it was necessary to find means of identifying the split-off amino acids directly. The identification of the amino acid after its regeneration by hydrolysis of the PTH is not entirely satisfactory as certain amino acids are not returned unmodified. A direct identification of the PTH's was therefore made the objective. This can be achieved by paper chromatography (J. Sjöquist, 1953; P. Edman and J. Sjöquist, 1956) for most of the PTH's, but PTH-arginine, PTH-histidine, and PTH-cysteic acid are preferably identified by paper electrophoresis. The iodine-azide reaction used for locating the spots has the weakness of being nonstoichiometric and rather unspecific. This method of location is therefore always supplemented by a method based on the very high UV absorption of the PTH's. For this purpose, I have found it convenient to use a fluorescent screen excited by UV light and having its secondary emission in the visible red (P. Edman and J. Sjöquist, 1956). The arrangement is shown in FIGURE 1. The spots appear as dark shadows on a bright red background. This allows the spots to be cut out from the paper, eluted, and the content of PTH determined. In conjunction with a method for quantitative microsynthesis of PTH's from free amino acids (J. Sjöquist, 1957*a*), this paper chromatographic procedure offers a quick and fairly accurate means of analyzing amino acid mixtures. In this connection, I also draw attention to the highly accurate and fully automatic procedure of Sjöquist (1957*b*) used for the same purpose. A partition column procedure is employed to separate the PTH's, and their UV absorption is automatically recorded in the effluent as they emerge. The application of the method to an hydrolysate of bovine serum albumin is indicated in FIGURE 2 and TABLE 1.

Another problem in the degradation of long sequences stems from the fact that the splitting off of the *N*-terminal amino acid requires a strongly acid medium. When aqueous acids are used this leads to a partial hydrolysis along the peptide chain with the consequent generation of new *N*-terminal amino acids; the results, therefore, become ambiguous after only a few degradative steps. However, the cleavage reaction is *nonhydrolytic* in nature. This fact

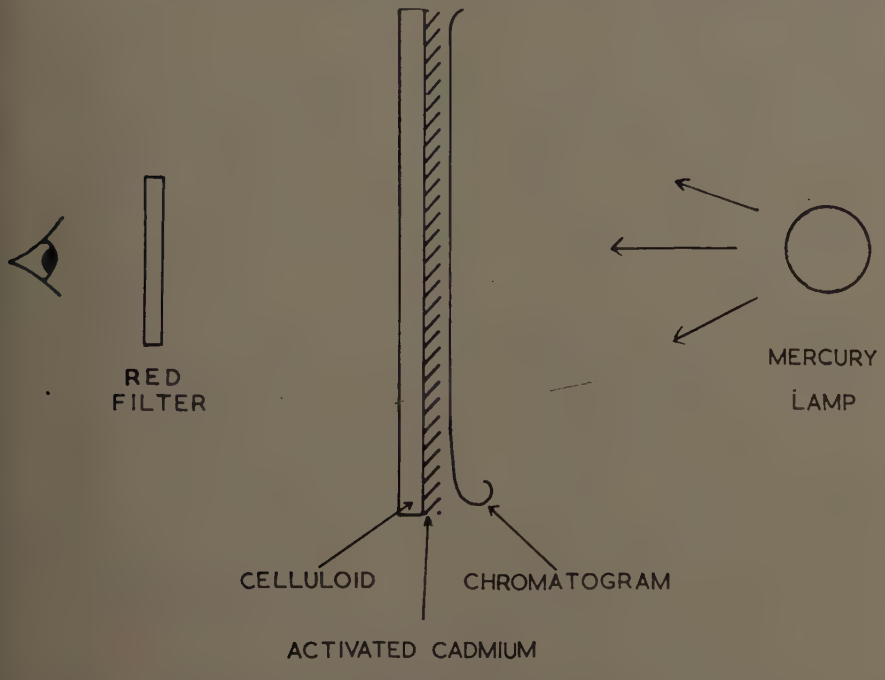


FIGURE 1. Set-up for locating UV-absorbing spots on paper chromatograms.

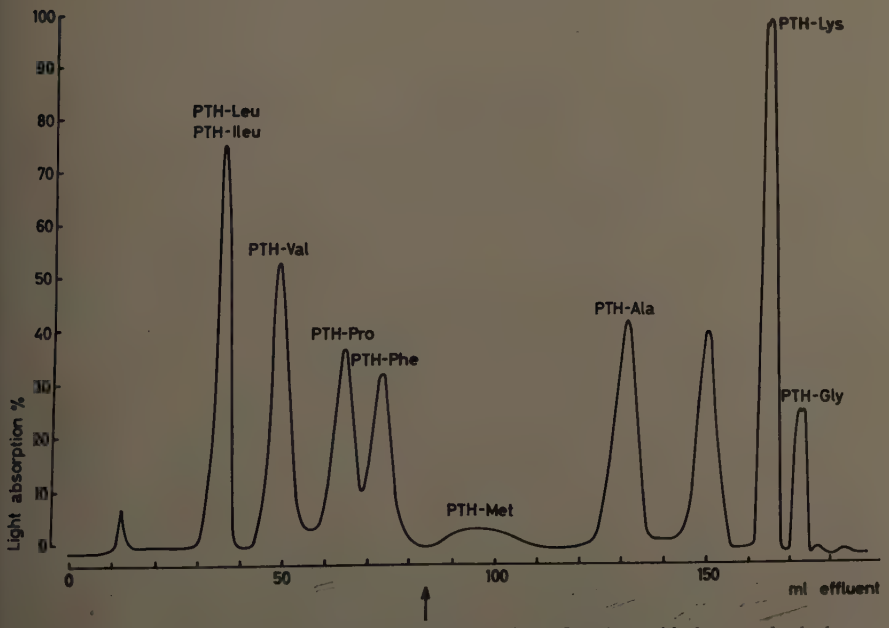


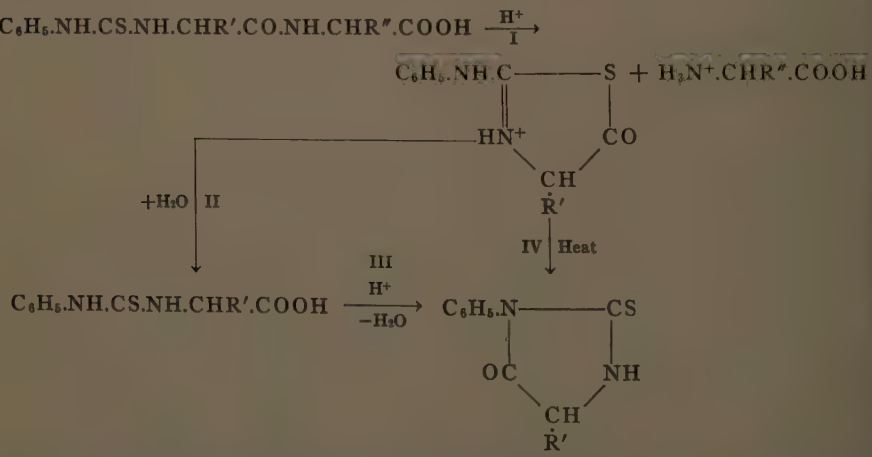
FIGURE 2. Column chromatographic separation of PTH-amino acids from an hydrolysate of serum albumin (J. Sjöquist, 1957*b*). Reproduced by permission from *Arkiv för Kemi*.

TABLE 1
AMINO ACID COMPOSITION OF BOVINE SERUM ALBUMIN

Amino acid	Amino acid/100 gm. protein			
	Sjöquist (1957*b*)		Stein and Moore (1949)	
	Grams	mM	Grams	mM
Leucines	14.75*	112.4	14.88	113.4
Valine	6.05	51.6	5.92	50.5
Proline	4.70	40.8	4.75	41.3
Phenylalanine	6.59	39.9	6.59	39.9
Methionine	0.68	4.6	0.81	5.4
Alanine	6.34	71.2	6.25	70.1
Lysine	12.31	84.2	12.82	87.7
Glycine	1.98	26.4	1.82	24.2
Tyrosine	—	150.0†	5.06	27.9
Glutamic acid	—		16.50	112.2
Threonine	4.11‡	34.5	5.83§	49.0
Aspartic acid	10.50	78.9	10.91	82.0
Serine	1.42‡	13.5	4.23§	40.2
Arginine	6.20	35.6	5.90	33.9
Histidine	4.15	26.7	4.00	25.8

* Calculated as leucine.
† Calculated as glutamic acid.
‡ Uncorrected for losses.
§ Corrected for losses.

has been clear from the inception of our work, and has been further emphasized by the clarification of the detailed mechanism of the cleavage reaction (P. Edman, 1956). This study has revealed that the *N*-terminal amino acid is split off as a 5-thiazolinone:



This derivative is unstable and can be converted by heat or by aqueous acid to the isomeric and stable 3-phenyl-2-thiohydantoin. It is important that Reaction I proceed in anhydrous acids and is fast (as is Reaction II), whereas Reactions III and IV are comparatively slow.

The importance of these observations bearing on the degradation procedure is that the exposure of the peptide or protein to anhydrous acid can be limited to the short time necessary for the completion of Reaction I. The 5-thiazolinone can then immediately be separated from the residual peptide or protein and converted in a separate operation to the isomeric PTH.

A cycle in the degradation will thus comprise three separate operations: (1) coupling of the peptide or protein with phenyl isothiocyanate to form the phenyl thiocarbamyl (PTC) derivative; (2) cleavage of the PTC-derivative in anhydrous acid into 5-thiazolinone and residual peptide; and, after removal of the latter, (3) conversion of the 5-thiazolinone to the corresponding PTH.

The application of the stepwise degradation directly to whole proteins has also been limited by the fact that, as a rule, proteins become insoluble and thus unreactive after a few degradation cycles. However, this difficulty is completely obviated when the disulfide bonds of the protein are first oxidized by performic acid. The insolubility of the unoxidized protein could be due to a polymerization of the protein as a consequence of disulfide interchange brought about by the anhydrous acid, but we have no direct evidence for this interpreta-

1	2	3	4	5	6		7	8	9	10	11		12		13	14
H.Lys.	Val.	Phe.	Gly.	Arg.	Cy	SO_3H.Glu.	Leu.	Ala.	Ala.	Ala.	MetO_2.		Lys.		Arg.	
H.Lys.	Val.	Phe.	Gly.	Arg.												
					.Cy	SO_3H.Glu.	Ala.	Leu.	Ala.	Ala.	MetO_2.		Lys.		Arg.	

FIGURE 3. The *N*-terminal sequence of oxidized lysozyme. The top row is the result of the present work, and the two lower rows are from the work of P. Jollès *et al.* (1958).

tion. In any case, a cleavage of the disulfide bonds is a necessary preliminary to a sequence determination, and thus fits naturally into the analytical scheme.

The *N*-terminal sequence of lysozyme has been determined in accordance with the outlined procedure (P. Edman, unpublished observation). Lysozyme from egg white was first oxidized with performic acid according to the procedure of J. G. Pierce (1955). Incidentally, it was found that the milder conditions used for the oxidation of ribonuclease (C. H. W. Hirs, 1956) did not bring about complete oxidation of the disulfide bonds of lysozyme. The PTH-amino acids were identified, as a rule, by paper chromatography in the three different solvent systems, with the exception of PTH-arginine and PTH-cysteic acid, which were identified by paper electrophoresis. The location of all spots was confirmed by their UV absorption using the fluorescent screen as described above, and their outlines marked in pencil on the chromatogram prior to the spraying with the iodine-azide reagent.

The sequence of 14 amino acids that has emerged from this study is presented in FIGURE 3 together with 2 related peptide fragments, the structures of which have been elucidated earlier (P. Jollès *et al.*, 1958) but not linked together. The sequences agree with one exception. The amino acids in position 8 and 9 occur in the reverse order in the structure of the French workers. We believe our sequence to be the correct one, as it has resulted from a straightforward degradation (FIGURE 4). Furthermore, the alternative of a glutamine residue

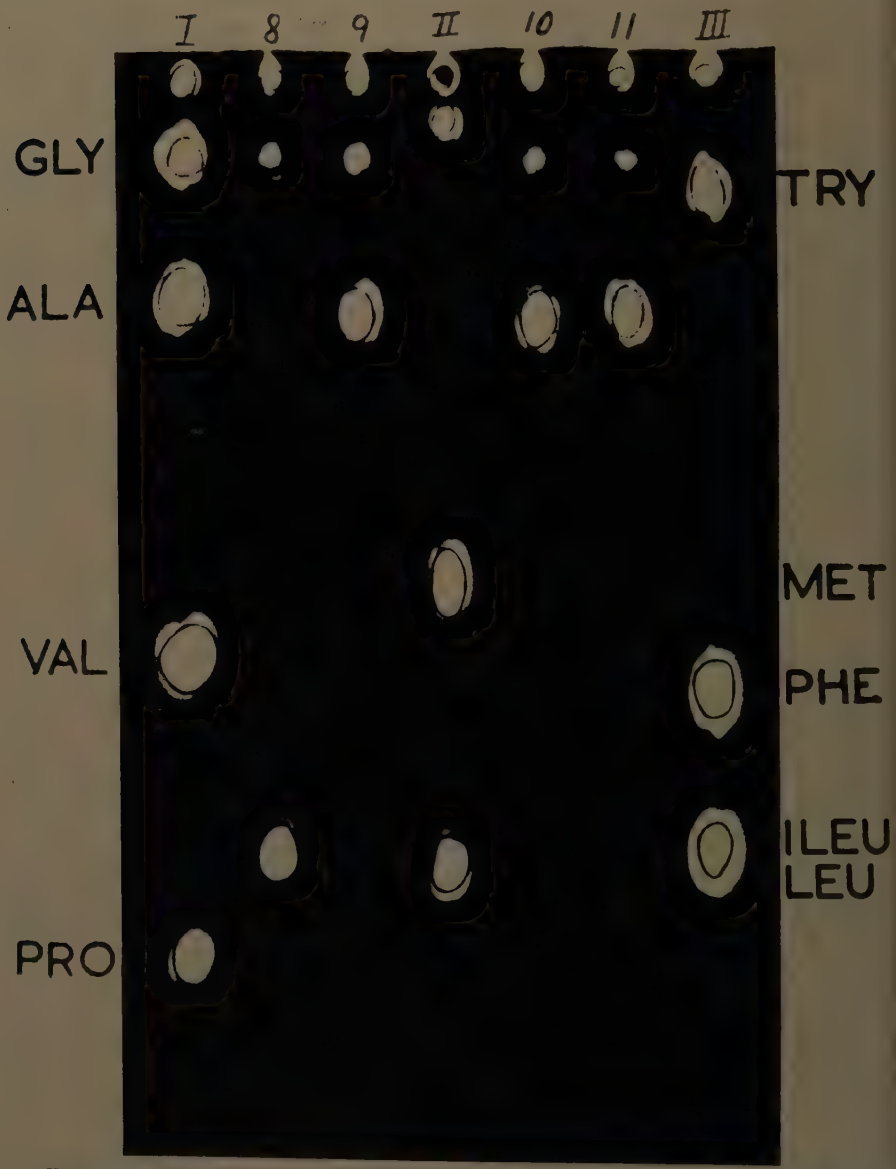


FIGURE 4. Paper chromatogram of the products from degradation cycles eight to eleven of lysozyme. Reference mixtures of PTH-amino acids are marked I, II, and III. UV-absorbing spots are outlined.

in position 7 is ruled out by my determination.* The degradation rather suddenly stopped to give clear results after the fourteenth cycles. It is con-

* The identification of lysine in position 13 is tentative, as it rests only on the somewhat uncharacteristic behavior of PTH-lysine in paper chromatography. A rigorous identification would have required, in addition, the demonstration of lysine in an hydrolysate of the PTH, but lack of material did not permit this confirmation.

ceivable that we encountered some irregularity at that stage, for example, an oxidation product of tryptophan.

Some remarks pertaining to the degradation method itself should be made:

(1) No overlap between consecutive steps was observed.

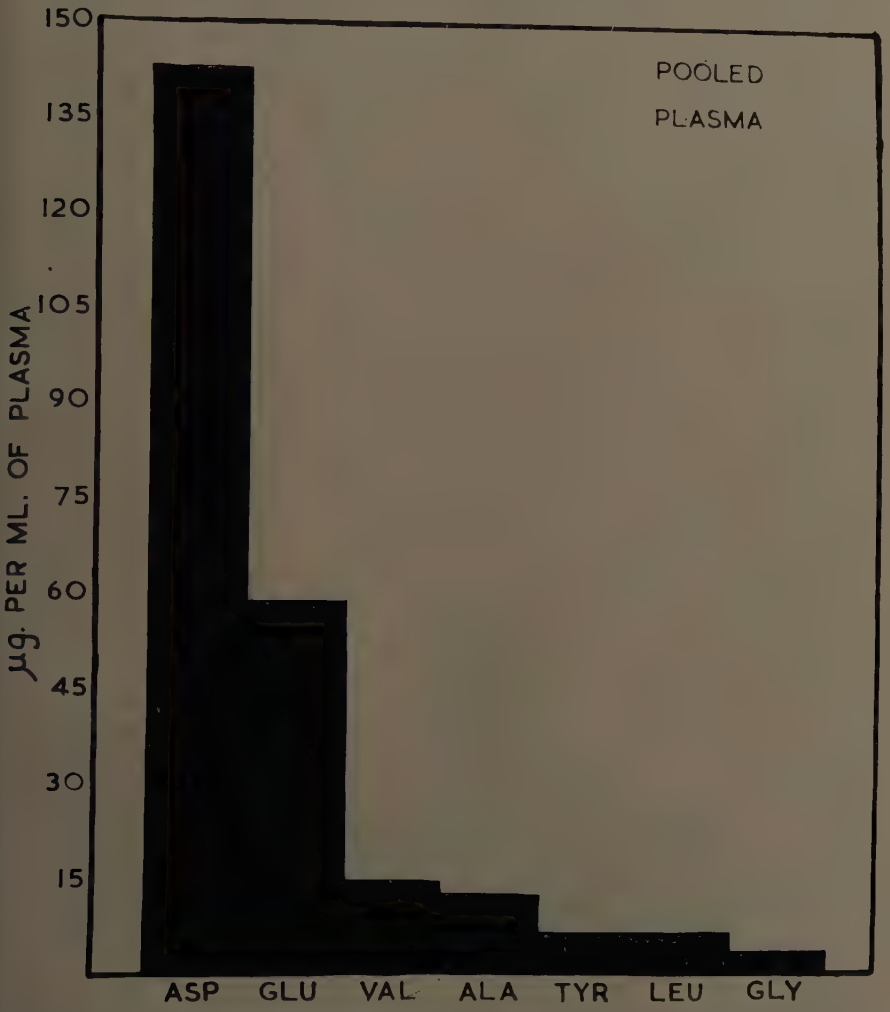


FIGURE 5. Distribution of *N*-terminal amino acids in the plasma proteins of healthy individuals. Quantities are in terms of PTH's.

(2) The average repetitive yield, that is, the yield from one step to the next, was approximately 90 per cent. About one half of the 10 per cent deficit could be accounted for by loss of protein during manipulation, and the over-all "chemical" yield in the coupling and cleavage reaction should therefore be over 90 per cent.

(3) In every step, traces of PTH-glycine, PTH-alanine, and what seems to be PTH-asparagine appeared (FIGURE 4). This confirms earlier observations

(G. Schramm *et al.*, 1956). The contaminants were present in too small amounts to cause doubt about the true sequence. As presumably they are derived from the cleavage of peptide bonds along the chain, it would be expected that the quantities of these by-products should be in some proportion to the

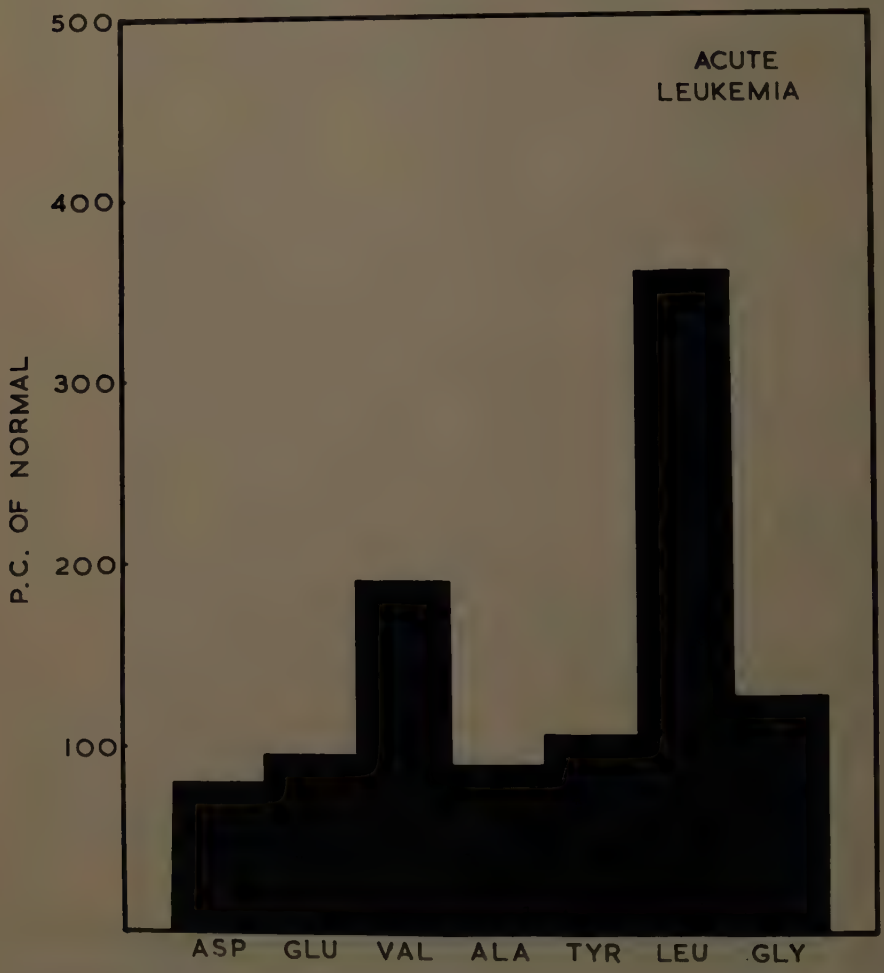


FIGURE 6*a*. Distribution of *N*-terminal amino acids in plasma proteins relative to normal in a case of acute leukaemia.

frequency of these amino acids in the sequence studied. In general, a larger structure should produce relatively more of them.

Before leaving the subject of sequence determinations I must stress the importance of the time factor in this kind of work. One degradation cycle can be completed in a normal working day, and it is entirely possible for one person to run two or three degradations at the same time. It should therefore now

be feasible to complete the sequence determinations of a structure the size of lysozyme in a few months provided the structure can be broken up into a few large fragments.

Finally I draw attention to a new field of application for *N*-terminal determinations by the PTH method. The rationale of this application is that a

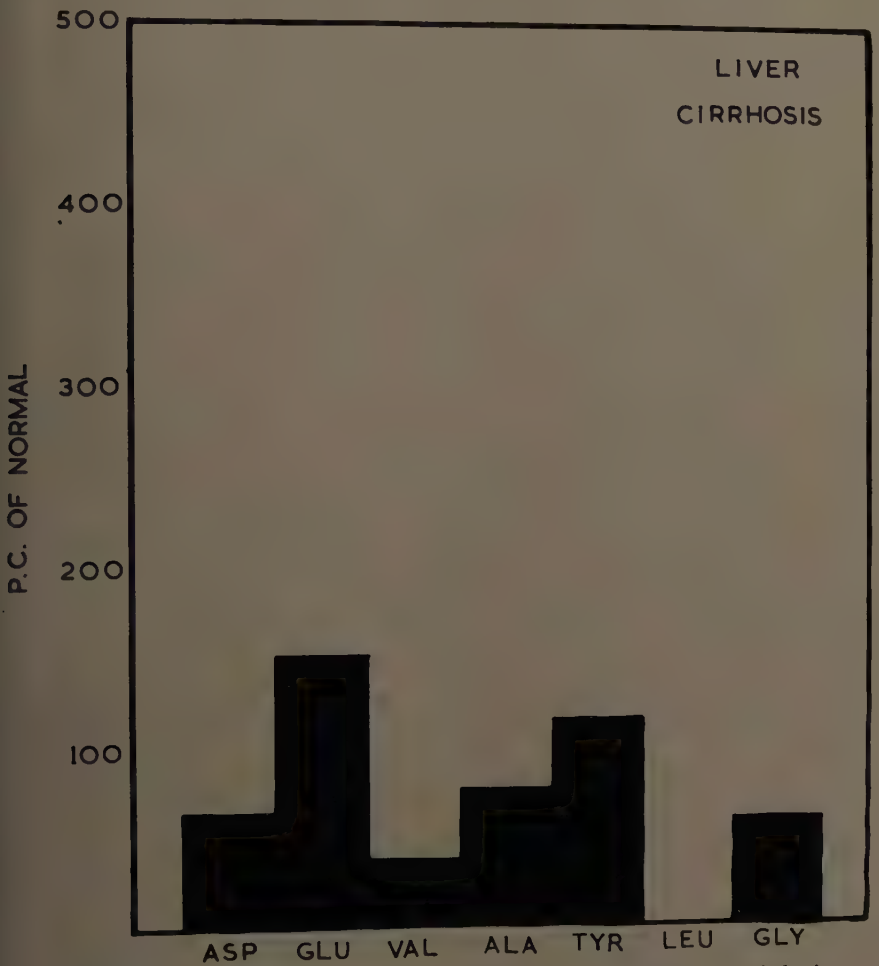


FIGURE 6*b*. Distribution of *N*-terminal amino acids in a case of liver cirrhosis.

mixture of proteins or peptides can be characterized by the qualitative and quantitative composition of its *N*-terminal amino acids. To date such a characterization has been possible only by physicochemical methods, for example, ultracentrifugation or electrophoresis. However, besides being simple to accomplish, *N*-terminal determinations have the advantage of yielding information about a fundamental property of the components that is independent of a certain set of experimental conditions. This approach could be

expected to yield interesting information when applied to physiological mixtures, for example, plasma proteins. We therefore have initiated an investigation of the *N*-terminal amino acid pattern of plasma proteins under normal and pathological conditions (H. Niall and P. Edman, unpublished observation), and I now indicate briefly some results. The reservation must be made that they are of a preliminary nature and may be open to future corrections in detail. Nevertheless it is already clear that radical changes in the *N*-terminal amino acid pattern do occur in pathological conditions. FIGURE 5 shows the distribution of *N*-terminal amino acids in healthy individuals. Our present technique allows us to identify seven amino acids: aspartic acid, glutamic acid, valine, alanine, tyrosine, leucine (isoleucine), and glycine. Aspartic acid dominates the pattern, as could be expected from the known fact that it occupies the *N*-terminal position in serum albumin (H. Van Vunakis and E. Brand, 1951). Examples of patterns obtained in pathological conditions are shown in FIGURE 6. Here the representation along the ordinate is shown in percentage of the normal in order to bring out clearly the great relative changes that can occur in the concentration of individual *N*-terminal amino acids, in this case leucine and valine.

References

EDMAN, P. 1949. Arch. Biochem. **22:** 475.
EDMAN, P. 1950. Acta Chem. Scand. **4:** 283.
EDMAN, P. 1953. Acta Chem. Scand. **7:** 700.
EDMAN, P. 1956. Acta Chem. Scand. **10:** 761.
EDMAN, P. & J. SJÖQUIST. 1956. Acta Chem. Scand. **10:** 1507.
HIRS, C. H. W. 1956. J. Biol. Chem. **219:** 611.
JOLLÈS, P., J. JOLLÈS-THAUREAUX & C. FROMAGEOT. 1958. Symposium on Protein Structure. : 277. A. Neuberger, Ed. Methuen. London, England.
PIERCE, J. G. 1955. J. Am. Chem. Soc. **77:** 184.
SCHRAMM, G., J. W. SCHNEIDER & A. ANDERER. 1956. Z. Naturforsch. **11b:** 12.
SJÖQUIST, J. 1953. Acta Chem. Scand. **7:** 447.
SJÖQUIST, J. 1957*a*. Arkiv. Kemi. **11:** 129.
SJÖQUIST, J. 1957*b*. Arkiv. Kemi. **11:** 151.
STEIN, W. H. & S. MOORE. 1949. J. Biol. Chem. **178:** 79.
VAN VUNAKIS, H. & E. BRAND. 1951. Abstracts of Papers, 119th Meeting Am. Chem. Soc. **28c.**

THE STRUCTURE OF RIBONUCLEASE

C. H. W. Hirs
Biology Department, Brookhaven National Laboratory, Upton, N.Y.

Introduction

With the passage of years, structural work with pancreatic ribonuclease has been carried out by an increasingly large group of investigators. It will be impossible in the space available here to be particularly comprehensive in dealing with so much work or, for that matter, fully to do justice even to those aspects with which I am most familiar. The present is nevertheless an appropriate time to attempt a brief summary of the current status of the problem posed by the structure of ribonuclease; from this summary it may be possible to discern certain trends in future work.

It is appropriate to begin this recapitulation by reminding ourselves that the advances now being made in the chemical investigation of protein structure are the logical outgrowth of the fundamental studies made almost twenty years ago in the laboratories of A. J. P. Martin and R. L. M. Synge, and of F. Sanger, in Cambridge, England. The original and far-ranging investigations of these workers were responsible for the introduction of what by now have become standard methods of approach and procedure in the polypeptide field. It was the dramatic success, attained through the application of these methods by Sanger and his colleagues, particularly H. Tuppy and E. O. P. Thompson,[1,2] in the elucidation of the amino acid sequences of the polypeptide chains in insulin, that stimulated and encouraged others to undertake structural studies with larger polypeptides. Particularly noteworthy successes have since then been attained with the establishment of the structures of a number of polypeptide hormones, notably the corticotrophins,[3,4,5] glucagon,[6] and the melanocyte stimulating hormones.[7,8,9]

When the present studies with ribonuclease were initiated in 1954 several considerations had influenced the selection of the protein for structural work.

Ribonuclease had long been recognized as an enzyme of exceptional stability. Indeed it was this attribute that led to its discovery in pancreatin by Jones in 1918.[10]

The fundamental work of Kunitz[11] on the isolation of ribonuclease from acid extracts of bovine pancreas made this enzyme a readily accessible material with which to conduct structural studies.

Rothen[12] had demonstrated that the crystalline enzyme isolated by Kunitz possessed a molecular weight of approximately 14,000 or roughly twice the molecular weight of the insulin monomer. The choice of ribonuclease therefore represented a reasonable advance up the molecular weight scale of proteins.

The analytical studies of Brand and his associates[11] had established the fact that the crystalline enzyme, like insulin, does not contain tryptophan, an amino acid of exceptional lability to acids and oxidizing agents, which made its absence a desirable feature in structure work.

Ribonuclease was one of the first proteins to be chromatographed successfully.[13,14] A convenient procedure was therefore available for assessing the homogeneity of the material used for structural studies.

In contrast to the large polypeptide hormones then being studied, ribonuclease was an enzyme, certain features of the specificity of action of which were already well-defined because of the investigations of Schmidt,[15] Markham and Smith,[16] Volkin and Cohn,[17] Brown and Todd,[18] and Heppel.[19] One of the principal goals in structural investigations in the polypeptide field has been the accumulation of sufficient information to permit an eventual explanation of function in terms of structure. The realization of this goal may be a little closer with enzymes than it is with hormones.

Preliminary Work

Having selected ribonuclease on the basis of these considerations, our first step in the program was an inquiry into the homogeneity of the protein. The importance of starting with homogeneous material in an investigation of the structure of a macromolecule like that of RNase need not be stressed here.

Our earlier work[14] had demonstrated that RNase could be chromatographed without inactivation on the carboxylic resin IRC-50. Chromatography on IRC-50 columns had also revealed the presence in dilute acid extracts of bovine pancreas of two proteins with RNase activity, the most prevalent of which was designated as ribonuclease (RNA). RNase is generally the main constituent of the crystalline enzyme as prepared by the method of Kunitz and McDonald.[20] Chromatography of a sample of crystalline RNase on a preparative scale afforded chromatographically homogeneous RNase, which proved to be indistinguishable in amino acid composition and chromatographic behavior from a sample of RNase prepared by direct chromatography of an acid extract of pancreas. The chromatographic procedure was therefore potentially useful for the convenient preparation of RNase.

Almost all of the amino acid sequence studies were carried out on a commercial preparation of crystalline RNase with an RNase content of 93 per cent. Before embarking on an extensive program of degradative work, however, it was first established that the crystalline RNase gave the same series of tryptic hydrolysis products as did RNase.

At the time the work was started it was known that RNase behaved as a homogeneous substance when chromatographed either on IRC-50 or by the method of Martin and Porter[13] on kieselguhr columns in an aqueous ammonium sulfate-ethyl cellosolve system, with constant specific activity in the elution peaks. Subsequently it was found that RNase is homogeneous on zone electrophoresis,[21] CCD,[22] and ultracentrifugation (Yphantis, this publication).

Upon quantitative amino acid analysis[23] as shown in TABLE 1, molar ratios for the constituent amino acids were obtained that closely approximated whole numbers, especially for those residues occurring in relatively small numbers in the protein, such as histidine, arginine, glycine, valine, leucine, and phenylalanine. The amino acid analysis in TABLE 1 revealed RNase to have the empirical formula $Asp_{16}.Glu_{12}.Gly_3.Ala_{12}.Val_9.Ileu_3.Leu_2.\frac{1}{2}Cys_8.Met_4.Ser_{15}.Thr_{10}.Pro_5.Phe_3.Tyr_6.(CONH_2)_{17}$. This formula was later shown to be in error when it became possible to account for the whole molecule in terms of smaller and more readily analyzed peptides. It was then discovered that there are 15 rather than 16 aspartic acid residues, and 4 rather than 5 proline residues in

RNase. The analysis also showed that essentially all the weight and nitrogen of the protein were accounted for in terms of amino acids, while the sulfur content was due exclusively to cystine and methionine. There is no phosphorus in RNase.

The presence of a single amino-terminal lysine residue was proved by Anfinsen and his colleagues[24] by means of the fluorodinitrobenzene (FDNB) procedure of Sanger. These investigators also studied the action of carboxypeptidase on RNase and concluded that valine was the only carboxyl-terminal amino acid. Their work therefore demonstrated that RNase contains a single peptide chain. This demonstration and the inability to detect any free thiol groups

Table 1

The Amino Acid Composition of RNase as Determined Prior to Detailed Structural Studies[23]

Amino acid	Amino acid residues per 100 gm. protein (gm.)	No. as per cent of total no.	No. of residues
Aspartic acid	13.0	8.86	15.8
Glutamic acid	10.9	6.64	11.8
Glycine	1.25	1.72	3.05
Alanine	6.12	6.77	12.0
Valine	6.34	5.03	8.95
Leucine	1.74	1.21	2.15
Isoleucine	2.30	1.60	2.85
Serine	9.44	8.54	15.0
Threonine	7.56	5.88	10.45
Half cystine	5.95	4.58	8.15
Methionine	3.52	2.10	3.75
Proline	3.32	2.69	4.79
Phenylalanine	3.13	1.67	2.97
Tyrosine	6.84	3.30	5.87
Histidine	3.73	6.41	3.80
Lysine	9.22	11.3	10.0
Amide		9.58	17.0
Total	98.8	96.8	

led to the postulation[24] of a structure internally cross-linked by 4 disulfide bonds between pairs of the 8 half-cystine residues.

Partial Structural Formula

In the interests of simplifying the degradation problem, RNase was oxidized with performic acid at -5 to $-8°$ C. to effect scission of the disulfide bonds.[25] The protein was freed of chloride ion prior to oxidation to prevent the generation of free halogen from the peracid, and thus chlorotyrosine formation was avoided. Oxidation at temperatures above 0° C. resulted in degradation of tyrosine, presumably by attack of the phenolic ring.

Quantitative amino acid analyses of the oxidized protein showed that a reaction period of 2 hours was just sufficient for the complete oxidation of cystine to cysteic acid. The four methionine residues in RNase were found to have been completely converted to methionine sulfone in the product. With

the possible exception of serine and threonine, all the other amino acids in the protein were recovered in the expected quantities. The slight variation in the losses of the hydroxyamino acids that take place on acid hydrolysis of a protein prior to amino acid analysis made it difficult to decide whether a small but significant decrease in the recovery of serine and threonine was due to oxidation.

Performic acid-oxidized RNase was found to contain no new amino-terminal groups by the FDNB method,[25] from which it was evident that the peptide chain had not been cleaved during treatment with performic acid. As a result of oxidation the RNase molecule unfolds dramatically with profound effects on many of the properties of the protein. Of most immediate interest is the fact that oxidation renders the protein readily susceptible to the action of trypsin and chymotrypsin. By contrast, RNase is not attacked appreciably by these enzymes.

Before considering the further degradation of RNase by way of the performic acid-oxidized protein, it should be made clear that, while the analytical characterization of the oxidized protein was satisfactory, no homogeneity studies were performed on this material. A few attempts at chromatography of oxidized RNase were unsuccessful.

To achieve a specific cleavage of the peptide chain in performic acid-oxidized RNase, trypsin was used at *p*H 7 at a relative molar concentration of approximately 1:400. There are 10 lysine and 4 arginine residues in RNase. Therefore a maximum of 15 peptides could arise if cleavage were complete at the carbonyl bonds of each of these residues. There was a considerable variation in the rate at which the trypsin-sensitive bonds were hydrolyzed. As we shall see later, a lysyl-proline bond (residues 41 to 42) was not cleaved at a significant rate, while the amino-terminal lysine residue in the protein[24] was found to be removable only with difficulty. Thus, in actuality, 13 principal peptides (TABLE 2) were formed on tryptic hydrolysis of the oxidized protein.[26] Because the arginyl-cysteic acid bond at residues 39 to 40 is cleaved relatively slowly, one of the intermediates in the hydrolysis (Peptide O-Tryp 13 in TABLE 2) was found to persist in the hydrolysate long after the initial, rapidly transient intermediates had disappeared.

TABLE 2 shows that the series of peptides formed on tryptic hydrolysis of oxidized ribonuclease (trypsin series) range in size from dipeptides like serylarginine and aspartylarginine to a peptide designated O-Tryp 9,* which contains 22 amino acid residues, and thus approximates the glycyl chain of insulin in size.

For the fractionation of peptides with such widely diverse molecular weights chromatography on Dowex 50-X2† was found to be particularly advantageous because of the high chromatographic efficiency that was attainable: peaks on the elution curves corresponding to large peptides were never more than twice as broad as those corresponding to small peptides moving with the same relative degree of retention on the columns. When similar fractionations were

* The peptides isolated from RNase were given arbitrary designations based on the approximate order in which they were eluted upon chromatography and the enzyme used to effect hydrolysis, with the prefix O- added to indicate that the protein had been oxidized prior to enzymatic hydrolysis.

† Manufactured by Dow Chemical Company, Midland, Mich.

attempted on the more highly cross-linked resin Dowex 50-X4 only the smaller peptides were chromatographed efficiently; peptides of intermediate size appeared as widely spread-out zones, while the larger peptides chromatographed so poorly that they could not be detected as discrete zones.

Chromatography on Dowex 50-X2 was carried out with the resin either in the sodium or the ammonium cycle. For analytical scale exploratory work the use of sodium buffer salts in the eluents was advantageous because it permitted straightforward detection of peptide zones by the ninhydrin color reaction. For preparative work, however, the use of volatile ammonium formate or acetate buffers for elution proved preferable, notwithstanding the problem of

TABLE 2

PEPTIDES ISOLATED FROM TRYPTIC HYDROLYSATES OF PERFORMIC ACID-OXIDIZED RNASE*

Peptide	Empirical formula in terms of residues	Per cent yield after	
		3 hr.	20 hr.
O-Tryp 2	$Asp_3,Glu_2,Gly,Ileu,Ser_3Thr_3,-(CySO_3H)_2,MeSO_2,Tyr_2,Arg$	55	15
O-Tryp 4	$Asp_3,Glu_2,Ala_2,Ser_6,Thr,CySO_3H,-(MeSO_2)_3,Tyr,His,Lys$	50	45
O-Tryp 5	$Asp,Ala,Val,CySO_3H,Lys$	95	95
O-Tryp 6	Glu,Gly,Ser_2,Thr,Lys	90	100
O-Tryp 7	Asp,Arg	70	90
O-Tryp 8	Asp,Glu,Ala,Thr_2,Lys	100	100
O-Tryp 9	$Asp_2,Glu_3,Ala_2,Val_4,Leu,Ser_2,Thr,-(CySO_3H)_2,Pro,Phe,His,Lys_2$	30	50
O-Tryp 10	Glu,Ala_3,Thr,Lys_2	95	100
O-Tryp 11	Asp,Leu,Thr,Lys	100	95
O-Tryp 12	Ser,Arg	55	55
O-Tryp 13	$Asp_3,Glu_3,Ala_2,Val_4,Leu,Ser_2,Thr,-(CySO_3H)_2,Pro,Phe,His,Lys_2,Arg$		20
O-Tryp 14	$Asp,Ala,CySO_3H,Pro,Tyr_2,Lys$		85
O-Tryp 15	Glu,Phe,Arg	75	85
O-Tryp 16	$Asp,Glu_2,Gly,Ala_2,Val_4,Ileu_2,Ser,-CySO_3H,Pro_2,Phe,Tyr,His_2$		65

* The amino acid residues are designated by standard abbreviations.

applying the ninhydrin reaction for the detection of peptides in ammonia-containing solvents. Where preparative work was performed with columns operated with sodium buffer salts, ion-exchange desalting of the peptide solutions was required. Since most of the peptides were strongly absorbed on the acid form of Dowex 50-X2, the procedure developed for this purpose depended upon absorbing the peptide from the sodium buffer solution at *p*H 2. The sodium in the resin was subsequently displaced with an excess of ammonium formate at *p*H 3, after which the peptide could be eluted with concentrated ammonium acetate solution at *p*H 7 and obtained as a lyophilized powder following the volatilization of the buffer salts.

Quantitative amino acid analysis of the isolated peptide fractions by the method of Moore and Stein and of Spackman *et al.*[27,28] not only revealed the yields in which the peptides had been formed but served to monitor the progress

of the fractionations, and indicated with a high degree of accuracy when a peptide had been isolated in a homogeneous condition.

All the residues in oxidized ribonuclease were accounted for by the peptides listed in TABLE 2. Since these peptides were formed in yields ranging from 50 to 100 per cent per molecule of RNase, they were the logical point of departure for an attack on the amino acid sequence in the polypeptide chain. Some of the peptides, however, notably O-Tryp 2, O-Tryp 4, O-Tryp 9, and O-Tryp 16 were too large to be studied directly and had to be broken down into subsidiary fragments of a more manageable size. Means to accomplish such further fragmentation had to be found. Moreover, the order in which the peptides, formed by the action of trypsin, occurred along the peptide chain had to be known before the amino acid sequence of the parent protein molecule could be reconstructed. Both of these problems could be approached by studying the peptides formed when oxidized RNase was subjected to the action of chymotrypsin.

There are 6 tyrosine and 3 phenylalanine residues in RNase. Chymotryptic hydrolysis took place at the carbonyl bonds of all of these residues except the tyrosine residue at position 92. This particular tyrosine residue is bonded to proline at position 93.* In addition to the principal cleavages at the carbonyl bonds of 5 aromatic amino acid residues, chymotryptic hydrolysis took place at widely different rates at many other bonds in oxidized RNase. In spite of the complexity of the resulting mixture of peptides, chromatography on Dowex 50-X2 permitted the separation and isolation[29] of all the major components of the mixture.

The chymotryptic hydrolysis gave a series of peptides (chymotrypsin series) in yields between 6 and 70 per cent which, as TABLE 3 shows, between them contained more than the total number of residues in RNase. Several sequences were thus being accounted for more than once by these peptides.

An independent, though less complete, series of peptides was obtained by Bailey[30] by the action of pepsin on oxidized RNase (pepsin series). These peptides are listed in TABLE 4.

Proceeding from the amino acid composition of RNase and of the peptides formed by the action of trypsin, chymotrypsin, and pepsin, the next step was the unambiguous identification of peptides of the three series that possessed residues in common. To do this it was necessary to make the reasonable assumption that trypsin catalyzes hydrolysis at the carbonyl bonds of lysine and arginine residues, while chymotrypsin catalyzes hydrolysis at the carbonyl bonds of phenylalanine and tyrosine residues. With these assumptions, the validity of which was to be proved later in any event, the peptides of the three series could be fitted together into a unique scheme as shown in FIGURE 1.[29] Since the sequence of residues in oxidized RNase is broken down into clusters by the peptide assignments depicted in FIGURE 1, such a scheme may properly be called a partial structural formula. This formula, while not providing the detailed information available from the complete sequence, is useful for an approximate delineation of the distribution of residues in the chain.

* Resistance of a tyrosylproline bond to chymotryptic hydrolysis and of a lysylproline bond to tryptic hydrolysis (compare residues 92–93 and 41–42, respectively, in FIGURE 5) was encountered in work on the amino acid sequence of the corticotrophins.[3]

TABLE 3

PEPTIDES ISOLATED FROM CHYMOTRYPTIC HYDROLYSATES OF PERFORMIC ACID-OXIDIZED RNASE*

Peptide	Empirical formula in terms of residues	Per cent yield after 24 hr.
O-Chy 2	Glu,Ser,Tyr	85
O-Chy 3	Asp,Glu,Gly,Ala,Val,$Ileu_2$,$CySO_3H$,Pro,Tyr	70
O-Chy 4	Asp,Ala,Val,Ser	60
O-Chy 5	Asp_2,Glu,Gly,Ala,Val,Thr,$(CySO_3H)_2$,Tyr,Lys	40
O-Chy 6	Ser,Thr,$MeSO_2$	80
O-Chy 11	Asp,Ala_2,Ser_3,Tyr	50
O-Chy 12	Asp_4,Glu_4,Gly,Ala_3,Val_4,Leu,Ser_2,Thr,$(CySO_3H)_3$,Tyr,His,-Lys_2	6
O-Chy 14	Asp_2,Glu,Gly,Ala,Ileu,Ser_3,Thr_2,$(CySO_3H)_2$,Pro,Tyr_2,Lys,Arg	55
O-Chy 15	Asp_2,Glu_3,Ala_2,Val_3,Leu,Ser_2,$CySO_3H$,His,Lys	30†
O-Chy 16	Asp_2,Glu_3,Ala_2,Val_3,Leu,Ser_2,$CySO_3H$,His,Lys	
O-Chy 19	Asp,Glu,Ala,Thr_2,Lys	26
O-Chy 21	Asp,Glu_2,Ser_3,Thr,$MeSO_2$,His,Arg	35
O-Chy 22	Asp_2,Glu_2,Ala_2,Ser_6,Thr,Tyr,His,Arg	25
O-Chy 25	Glu,Ala_3,Thr,Phe,Lys_2	75
O-Chy 27	Asp_2,Glu,Leu,Ser,$CySO_3H$,$(MeSO_2)_2$,Lys,Arg	50
O-Chy 29	Asp,Glu,Gly,Ala,Val,$Ileu_2$,$CySO_3H$,Pro,Tyr,His,Lys	35
O-Chy 30	Val_2,Pro,Phe,His	75
O-Chy 31	Asp_2,Val,Thr_2,$CySO_3H$,Pro,Phe,Lys_2,Arg	65

* The amino acid residues are designated by standard abbreviations.
† Combined yield of O-Chy 15 and O-Chy 16 (cf. Hirs *et al.*[29]).

TABLE 4

PEPTIDES ISOLATED FROM PEPTIC HYDROLYSATES OF PERFORMIC ACID-OXIDIZED RNASE*

Peptide	Empirical formula in terms of residues	Per cent yield after	
		4 hr.	24 hr.
O-Pep 2	Asp,Ala,Val,Ser	95	65
O-Pep 3	Asp,Glu,Ala,Val	95	83
O-Pep 5	Asp,Glu,Gly,Ala,Val_2,$CySO_3H$,Pro,Phe,Tyr,His	10	41
O-Pep 6	Glu,Ala_2,Thr,Lys	20	100
O-Pep 7	Glu,Val,Leu,Ser,His	10	29
O-Pep 9	Glu,Ala_3,Thr,Phe,Lys_2	52	0
O-Pep 10	Glu,Val,Leu,Ser,Phe,His	46	76
O-Pep 11	Asp_6,Glu_4,Ala_2,Val_2,Leu_2,Ser_3,Thr_3,$(CySO_3H)_2$,-$(MeSO_2)_3$,Pro,Phe,Tyr,His_2,Lys_3,Arg_3	43	

* The amino acid residues are designated by standard abbreviations.

The derivation of this formula may be approached in several ways, of which the most rigorous is the one that gives the main emphasis to the chymotryptic peptides obtained in the highest yields.*

* The derivation of this formula may be approached in several ways, all of which ultimately lead to the same unique conclusion. The derivation given here differs from that presented elsewhere[30] in that no information on amino acid sequences or terminal groups is utilized, and is presented because the importance of FIGURE 1 in the subsequent argument makes the demonstration of an alternative deduction desirable.

Derivation of Partial Structure

Section 1. Considering the tyrosine residues first, it will be observed that peptides O-Chy 11, O-Chy 2, and O-Chy 3 do not contain lysine or arginine and therefore come from parts of the sequence that do not correspond to overlaps between adjacent peptides of the trypsin series.

Peptide O-Chy 11 contains two alanine residues and cannot arise from the sequence covered by O-Tryp 2 or 14, both of which have less than two residues

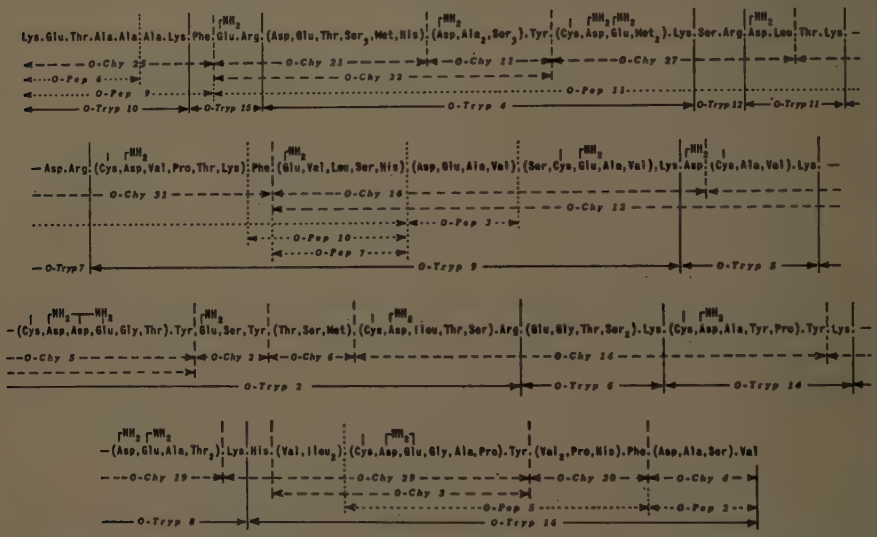


FIGURE 1. A partial structural formula for oxidized ribonuclease.[29] For presentation, the single peptide chain has been broken into four segments. Amino acid residues in parentheses are in undetermined sequences. Methionine sulfone and cysteic acid residues are represented as Met and Cys, respectively. The peptides obtained from tryptic hydrolysates are shown by the solid lines, those from chymotryptic hydrolysates by the dashed lines, and those from peptic hydrolysates by the dotted lines. The locations of the amide groups shown here were those originally derived from approximate determinations of amide ammonia of the isolated peptides, and comparison with FIGURE 5 will reveal that a number of these allocations were subsequently found to have been in error. Reproduced by permission from the *Journal of Biological Chemistry*.

of alanine. Peptide O-Chy 11 also contains three serine residues and thus could not arise from the part of the sequence covered by O-Tryp 16, which contains but one serine residue. O-Chy 11 is therefore left, by elimination, to come from only one place, namely, the same sequence as is covered by O-Tryp 4, and we may conclude that O-Tryp 4 and O-Chy 11 contain the same tyrosine residue.

Peptide O-Chy 3 contains two isoleucine residues and therefore arises from the same sequence as O-Tryp 16, the only peptide of the trypsin series with two isoleucine residues. Peptide O-Chy 29 contains the same residues as O-Chy 3, notably the two isoleucine residues just mentioned, with the addition of one residue each of histidine and lysine. O-Chy 29 is therefore an overlap peptide,

and this reveals that O-Chy 29 and O-Chy 3 are derived from the amino-terminal portion of O-Tryp 16.

Peptide O-Chy 2 contains serine and therefore comes from the same part of the sequence as O-Tryp 2, but not O-Tryp 14, which contains no serine.

Section 2. The remaining three tyrosine residues in peptides of the chymotrypsin series are present in O-Chy 5 (one residue) and O-Chy 14 (two residues). Both peptides contain basic amino acid residues and therefore correspond to overlaps of peptides of the trypsin series.

Peptide O-Chy 5 contains valine, glycine, and lysine. Valine is present in peptides O-Tryp 9 (4 residues), O-Tryp 5 (1 residue), and O-Tryp 16 (4 residues). O-Tryp 16 contains no basic amino acids and is therefore the carboxyl-terminal trypsin-series peptide. Thus O-Chy 5 represents an overlap of the sequences of either O-Tryp 9 or 5 and a tyrosine-containing peptide of the trypsin series that also contains glycine. There is only one such peptide, O-Tryp 2. O-Tryp 2 therefore follows either O-Tryp 9 or O-Tryp 5 in the trypsin series.

The remaining two tyrosine residues are present in O-Chy 14, a peptide that contains both arginine and lysine and thus represents an overlap that includes a sequence covered by three adjacent trypsin series peptides. One of these is O-Tryp 14, which contains the remaining two unassigned tyrosine residues in the trypsin series. O-Chy 14 also contains glycine, isoleucine, and proline. The isoleucine residue identifies O-Tryp 2, the only peptide outside of O-Tryp 16 in the trypsin series that contains this amino acid. Since the glycine residue in O-Chy 5 is the same as that in O-Tryp 2, the glycine residue in O-Chy 14 must come from a sequence other than that of O-Tryp 2. Only two peptides come into consideration, O-Tryp 6 and O-Tryp 16. The latter is ruled out because O-Chy 3 accounts for the glycine residue in O-Tryp 16, which in any event is the carboxyl terminal peptide and cannot come ahead of O-Tryp 14 in the sequence. This leaves O-Tryp 6 to contain the same glycine residue as O-Chy 14. The arginine residue in O-Chy 14 therefore corresponds to the arginine of O-Tryp 2 and the lysine residue to the lysine in O-Tryp 6. This assignment thus extends the trypsin series to include the following: O-Tryp 9 or 5—O-Tryp 2—O-Tryp 6—O-Tryp 14.

Section 3. The assignments of the last paragraph have covered the carboxyl and amino ends of O-Tryp 2. It may now be observed that tyrosine peptide O-Chy 22 and peptide O-Chy 21 contain arginine, methionine sulfone, and histidine, which identify these peptides as arising in part from O-Tryp 4 and an arginine peptide of the trypsin series. (Methionine sulfone and histidine are simultaneously present only in O-Tryp 4.) It is also to be noted that the amino acid composition of O-Chy 22 is the sum of the amino acid compositions of O-Chy 11 (compare with paragraph 3, *Section 1* above) and O-Chy 21. The three peptides O-Chy 11, 21, and 22 account for the amino end of O-Tryp 4, and it may be inferred that O-Chy 22, originally formed in high yield (approximately 75 per cent), is further hydrolyzed by secondary chymotryptic cleavage into O-Chy 21 and O-Chy 11. Since the arginine residue of O-Tryp 2 has already been allocated, the arginine residue of O-Chy 22 corresponds to the arginine of either O-Tryp 7, 12, or 15. O-Tryp 7 (aspartylarginine) and O-Tryp

12 (serylarginine) are dipeptides and, if they contained the same arginine residue as O-Chy 22, the latter would have to be formed by a primary chymotryptic hydrolysis at either lysine, arginine, serine, or aspartic acid, all of which are unlikely to be the sites of effective cleavage by this enzyme. From this reasoning it may be concluded that O-Chy 22 contains the same arginine residue as O-Tryp 15, which therefore identifies the sequence O-Tryp 15—O-Tryp 4. Space does not permit a further demonstration that the alternative assignments lead to untenable conclusions in the subsequent argument.

Section 4. Of the three phenylalanine-containing peptides of the chymotrypsin series O-Chy 25, O-Chy 30, and O-Chy 31, O-Chy 30 does not contain a basic amino acid and therefore does not represent an overlap between the sequences of two adjacent trypsin-series peptides. O-Chy 30 could be derived from the sequence of either O-Tryp 9 or O-Tryp 16. In the third paragraph of Section 1 it was shown that O-Chy 29 and O-Chy 3 represent the amino-terminal portion of O-Tryp 16. If O-Chy 31 contained the phenylalanine residue of O-Tryp 16, the presence of basic amino acids in O-Chy 31 would require that this peptide came from the amino-terminal portion of O-Tryp 16. Since this has already been shown to be explained by O-Chy 29 and O-Chy 3, peptide O-Chy 31 cannot contain the same phenylalanine residue as peptide O-Tryp 16. Thus O-Chy 30 contains the same phenylalanine residue as O-Tryp 16.

Since it is not possible for O-Chy 31 to come from the same sequence as O-Tryp 16, O-Chy 31 contains a sequence of residues involving an overlap of at least three trypsin series peptides, one of which is identified as O-Tryp 9 by the presence of phenylalanine and proline in O-Chy 31. O-Tryp 9 is the only remaining peptide of the trypsin series that could contribute these two residues. The presence of arginine in O-Chy 31 further reveals that the sequence covered by O-Chy 31 must involve the sequence of either O-Tryp 7 (aspartylarginine) or O-Tryp 12 (serylarginine). Since O-Chy 31 does not contain serine, however, the arginine peptide involved is O-Tryp 7 (aspartylarginine). Supporting this assignment is the composition of peptide O-Tryp 13, which has an amino acid content equal to the sum of the amino acid contents of O-Tryp 9 and O-Tryp 7. Since O-Chy 31 represents the amino-terminal portion of O-Tryp 9, these considerations reveal the sequence (lysine peptide)—O-Tryp 7—O-Tryp 9.

It has just been shown that the phenylalanine peptide pairs O-Tryp 9 and O-Chy 31 and O-Tryp 16 and O-Chy 30 contain common residues of phenylalanine. Since there are three phenylalanine residues in RNase, by difference peptides O-Tryp 15 and O-Chy 25 contain the same phenylalanine residue. Peptide O-Chy 25 has the same amino acid residues as O-Tryp 10 with the addition of one phenylalanine residue. The unique presence of three alanine residues in these peptides further identifies a common sequence. Thus the connection O-Tryp 10—O-Tryp 15—O-Tryp 4 may be written.

Section 5. Peptides of the chymotrypsin series devoid of phenylalanine and tyrosine must now be considered.

Of these, O-Chy 27 contains leucine, two residues of methionine sulfone, aspartic acid, lysine, and arginine, and represents an overlap covering the

sequences of parts of three trypsin peptides. The distribution of methionine sulfone permits an identification of these trypsin peptides. In the second and third paragraph of *Section 2*, above, it was seen that the overlap peptides corresponding to the amino- and carboxyl-terminal ends of O-Tryp 2 are O-Chy 5 and O-Chy 14, respectively. Both O-Chy 5 and O-Chy 14 do not contain methionine sulfone. The amino end of O-Tryp 4 is covered by O-Chy 21 and O-Chy 22, which between them account for one methionine sulfone residue. Since the amino and carboxyl ends of O-Tryp 2, which contains one methionine sulfone residue, are accounted for, and since O-Chy 27 is an overlap peptide, both methionine sulfone residues in O-Chy 27 are derived from the same sequence as O-Tryp 4. By difference, the remaining methionine sulfone residue in the chymotrypsin series, in peptide O-Chy 6, is seen to be the same as the methionine sulfone residue in O-Tryp 2. Since O-Chy 27 is seen to be derived from the carboxyl end of O-Tryp 4, the lysine residue in O-Chy 27 is the same as that in O-Tryp 4. The presence of arginine in O-Chy 27 demonstrates that an arginine peptide succeeds O-Tryp 4 in the trypsin series. Moreover, since the serine content of O-Tryp 4 is accounted for by the sequence covered by O-Chy 11 and O-Chy 21, the serine residue in O-Chy 27 must be derived from another sequence. Only one arginine containing peptide of the trypsin series is unassigned, O-Tryp 12 (serylarginine). It therefore follows that O-Tryp 12 succeeds O-Tryp 4 in the sequence of trypsin peptides and that the serine in O-Tryp 12 is the same as that in O-Chy 27. The presence of leucine in O-Chy 27 demonstrates that peptide O-Tryp 11 follows O-Tryp 12 in the trypsin series. The amino-terminal portion of O-Tryp 9, a peptide which also contains leucine, is covered by the sequence of O-Chy 31 (cf. second paragraph of *Section 4*, above) that does not contain leucine. The leucine residue in O-Tryp 9 therefore cannot be the same as the leucine in O-Chy 27.

Peptide O-Chy 16 contains the second of the two leucine residues in the parent molecule. The amino-terminal portion of O-Tryp 9 is covered by the sequence of O-Chy 31 (cf. second paragraph of *Section 4*, above). Therefore peptide O-Chy 16 contains the same leucine residue as O-Tryp 9, and this leucine residue must be somewhere on the carboxyl side of the phenylalanine residue in O-Tryp 9. Because O-Tryp 9 contains two lysine residues it is not possible to say whether O-Chy 16 is an overlap peptide on the basis of its composition alone. However, the large peptide O-Chy 12, with an amino acid content equal to the sum of the amino acid contents of O-Chy 16 and O-Chy 5, makes a decision on this point possible. Since peptide O-Chy 5 represents a portion of the sequence between either O-Tryp 9 or 5 and O-Tryp 2 (cf. second paragraph of *Section 2*, above) and since O-Tryp 2 contains no alanine and O-Tryp 9 contains two alanine residues, the presence of three alanine residues in O-Chy 12 demonstrates that an alanine-containing peptide is interposed between O-Tryp 9 and O-Tryp 2, a connection which is made obvious by the presence of four valine residues in O-Chy 12. Since O-Chy 5 contains one valine residue (as well as one lysine residue), the interposed peptide is a lysine-containing peptide that also contains a residue of alanine and a residue of valine. There is only one such peptide, O-Tryp 5. From this the sequence O-Tryp 9—O-Tryp 5—O-Tryp 2 follows.

Section 6. This leaves one peptide of the trypsin series to be allocated, O-Tryp 8. In the previous sections the following groups of trypsin peptides were connected: O-Tryp 10—O-Tryp 15—O-Tryp 4—O-Tryp 12—O-Tryp 11, O-Tryp 7—O-Tryp 9—O-Tryp 5—O-Tryp 2—O-Tryp 6—O-Tryp 14, and the carboxyl terminus, O-Tryp 16.

The presence of amino-terminal lysine in RNase requires that O-Tryp 10, a peptide that contains two lysine residues, be the amino-terminal peptide. O-Tryp 8 must therefore be assigned to a position between O-Tryp 11 and O-Tryp 7 or between O-Tryp 14 and O-Tryp 16. The decision between these two alternatives may be made with the help of the pepsin peptide O-Pep 11. This very large peptide contains two leucine and two histidine residues. It therefore embraces a sequence partly covered by O-Tryp 4 and O-Tryp 9, together with the peptides in between. The presence of three arginine residues in O-Pep 11 demonstrates that a portion of O-Tryp 15 is covered by the sequence of O-Pep 11, but the presence of two leucine residues requires that the single phenylalanine residue in O-Pep 11 be the same as that in O-Tryp 9. If O-Tryp 8 were derived from the sequence between O-Tryp 4 and O-Tryp 9, O-Pep 11 would have to contain 5 residues of threonine. In actuality, O-Pep 11 contains three residues of this amino acid, from which it is evident that O-Tryp 8 cannot be allocated to a position between O-Tryp 11 and O-Tryp 7. It is therefore placed between O-Tryp 14 and O-Tryp 16.

Section 7. From the sequence of the trypsin peptides derived in *Sections 1* to *6* above, the arrangement shown in FIGURE 1 is derived by inserting the chymotrypsin-series peptides from the amino-terminal end.

The relative alignment of O-Chy 2 and O-Chy 6 in the sequence covered by O-Tryp 2 could not be determined from the evidence presented thus far. In subsequent work, a peptide containing the first ten amino acid residues of O-Tryp 2 was isolated from oxidized RNase,[31] and the composition of this peptide unambiguously located O-Chy 2 ahead of O-Chy 6 in the sequence.

The allocation of peptide O-Chy 19 to the portion of the sequence covered by O-Tryp 8 was by difference. An interesting feature of the two peptides O-Chy 19 and O-Tryp 8 predicted by the partial structural formula is that they are isomers, involving the location of a lysine residue either at the amino- or carboxyl-terminal end of a common pentapeptide sequence. Subsequent structural work confirmed this prediction.

Peptide O-Chy 4 and O-Pep 2 contain the same amino acids and are placed at the carboxyl-terminal end of the molecule. The justification for this is provided by peptide O-Pep 5, which contains two proline residues, phenylalanine, and tyrosine, and established that peptides O-Chy 3 and O-Chy 30 are adjacent members of the chymotrypsin series.

The assignment of peptides O-Pep 7 and O-Pep 10 followed from the presence of leucine, histidine, and serine in both of these peptides, and the fact that these three amino acids are simultaneously present only in O-Tryp 9.

The assignment of peptide O-Pep 3 to the sequence covered by O-Tryp 9 rather than O-Tryp 16 was based on the yields of peptides O-Pep 3 and O-Pep 5, which showed that the two peptides could not be derived from the same sequence. The actual location of O-Pep 3 in FIGURE 1, however, was arbitrary.

Subsequent work demonstrated that the assignment of O-Pep 3 was actually the correct one.

The location of O-Pep 6 and O-Pep 9 was determined by their unique composition.

The peptide assignments given in the partial structural formula shown in FIGURE 1 were strengthened by the results obtained by Redfield and Anfinsen[32] through the application of a completely different approach. Oxidized RNase was dinitrophenylated with FDNB, and the resulting protein subjected to tryptic hydrolysis. On the assumption that all the ϵ-amino groups in the protein were alkylated by the reagent,* tryptic hydrolysis should be restricted to the carbonyl bonds of the four arginine residues. End-group analysis of the mixture of peptides formed by the action of trypsin by the FDNB method indeed gave results that indicated that four new amino-terminal residues were being produced, one aspartic acid, two glutamic acid, and one cysteic acid terminal group. The mixture of dinitrophenylated peptides proved to be particularly difficult to fractionate. Reference to FIGURE 1 will indicate that two of these fragments should have been relatively small: one of them a peptide corresponding to O-Tryp 10 plus O-Tryp 15, and another to peptide O-Tryp 11 plus O-Tryp 7. These peptides were indeed isolated, and the structure of the one corresponding to O-Tryp 11 plus O-Tryp 7 was established by partial acid hydrolysis. The isolation and characterization of the three larger fragments gave results that indicated that one peptide with a glutamic acid terminal group corresponded to the peptides from O-Tryp 6 to the carboxyl end of the chain; the largest peptide with a cysteic acid terminal group represented the sequences of peptides O-Tryp 9 through O-Tryp 2; and the second peptide with an amino terminal glutamic acid group corresponded to O-Tryp 4 plus O-Tryp 12. The relative alignment of these five peptide fragments was arrived at by studying the arginine-containing peptides formed upon partial acid hydrolysis and peptic hydrolysis of the dinitrophenylated protein.

The partial structural formula was checked by examining the products formed upon chymotryptic hydrolysis of the four largest peptides of the trypsin series, O-Tryp 4, O-Tryp 9, O-Tryp 2, and O-Tryp 16 (total number of residues in these peptides, 82).[31] The chymotryptic cleavages in FIGURE 1 should predict the manner in which these peptides are split by chymotryptic hydrolysis. How these predictions were born out is illustrated in FIGURE 2, which should be compared with FIGURE 1. The comparison shows that peptide O-Tryp 4 is cleaved by the enzyme as predicted by FIGURE 1. The action of chymotrypsin on O-Tryp 9 should lead to the formation of two subsidiary peptides according to FIGURE 1. At the higher enzyme concentrations used for the experiments summarized in FIGURE 2, the larger fragment arising from the carboxyl end of O-Tryp 9 was further broken down, principally by hydrolysis at two additional bonds. With O-Tryp 2, the action of the enzyme gave the products to be

* The dinitrophenylation of oxidized RNase was recently investigated in this laboratory (unpublished experiments). It proved to be impossible to achieve complete dinitrophenylation of all the ϵ-amino groups of lysine by employing the customary conditions for the reaction. The equivalent of one ϵ-amino group remained unsubstituted. A similar result was obtained with RNase itself. This observation is reminiscent of the behavior shown by RNase upon guanidination with O-methylisourea.[33]

expected from FIGURE 1, while with O-Tryp 16 breakdown by chymotrypsin was as anticipated from FIGURE 1, except that the amino-terminal histidine residue was removed extremely slowly.

The peptide assignments in FIGURE 1 were to receive further confirmation once the sequence of the amino acid residues in the peptides of the trypsin series had been worked out. These sequences predicted the identity of the

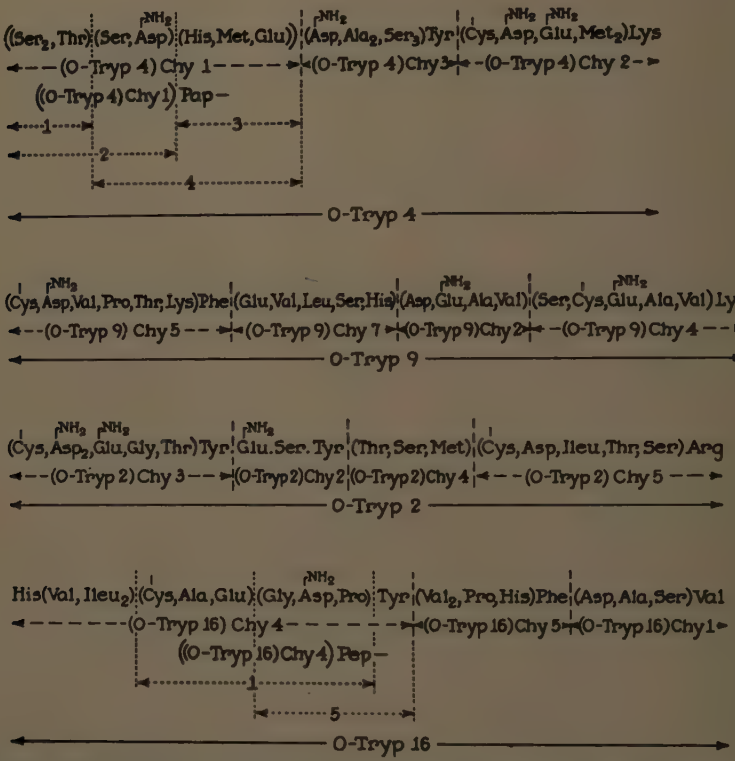


FIGURE 2. The subpeptides derived from peptides O-Tryp 4, 9, 2, and 16 by the action of chymotrypsin, pepsin, and papain.[31] The points of cleavage by chymotrypsin are shown by the dashed lines, while bonds attacked by pepsin in peptide (O-Tryp 16)Chy 4 and by papain in peptide (O-Tryp 4)Chy 1 are indicated by dotted lines. Reproduced by permission from the *Journal of Biological Chemistry*.

amino-terminal amino acids of the peptides of the chymotrypsin series. When the amino-terminal amino acids of the most important chymotrypsin series peptides were determined, the anticipated terminal amino acid was found in each peptide.

The Amino Acid Sequence

The knowledge summarized in FIGURE 2 was used to reduce the large trypsin series peptides, O-Tryp 4, O-Tryp 2, O-Tryp 9, and O-Tryp 16 to 14 subpeptides of a size more nearly suited to the application of methods for the determination of amino acid sequences.[31] The remaining 9 peptides of the trypsin series were

used without further degradation. The manner in which the total of 23 peptides that constituted the starting point of the sequence analysis were derived from oxidized ribonuclease is summarized in the chromatographic flow sheet shown in FIGURE 3.*

Sequence analysis was accomplished by a combination of chemical and enzymatic procedures. The main reliance was placed on the stepwise degradation of the peptides by the phenyl-*iso*-thiocyanate procedure of Edman.[34] As applied in the present work, the amino-terminal amino acid, cyclized off as the phenylthiohydantoin (or anilinothiazolinone) at each step of the procedure, was identified by quantitative amino acid analysis of the remaining peptide (difference analysis). This approach was occasionally supplemented by qualitative identification of the eliminated phenylthiohydantoin with the aid of paper chromatography. The phenyl-*iso*-thiocyanate procedure was also used in the same fashion for the straightforward identification of amino-terminal groups. The results of such end-group analysis were sometimes checked also by the FDNB procedure, in which event identification of the terminal amino acid was achieved by quantitative amino acid analysis of the residual peptide and by quantitative paper chromatography of the dinitrophenyl-amino acid fraction.

For the identification of carboxyl-terminal amino acids, particularly when these were either lysine or arginine, hydrazinolysis[35] was found to be an effective method. Less success attended efforts to identify some of the other carboxyl-terminal amino acids in the peptides studied by this method.

Quantitative measurements of the rate of release of free amino acids from peptides by the action of leucine aminopeptidase (LAP) or carboxypeptidase were used to obtain evidence for sequences at the same time that experiments on stepwise degradation were being performed. The successful application of enzymatic hydrolysis for sequence analysis depends upon a favorable relation between the relative rate of release of different amino acids on the one hand and the relative abundance as well as the sequence in which the amino acids are arranged on the other. In addition, in dealing with small peptide substrates the analytical problem of distinguishing between the released free amino acids and intermediate peptides assumes importance. The relative rate of release of the different amino acids, as measured with synthetic substrates (cf. Green and Neurath[36] and Smith and Spackman[37]), falls over a wide range and is also to some extent a function of the nature of the adjacent residues. As an example, the action of leucine aminopeptidase on the hexapeptide Ser.Ileu.Thr.-Asp.$CySO_3H$.Arg from oxidized RNase[38] may be considered. The rate of cleavage of the serylisoleucine bond is much slower than the rate at which the isoleucylthreonine bond is attacked. This disparity in the relative rates of hydrolysis of these bonds is reflected analytically by the simultaneous appearance in equal quantities of serine and isoleucine. The sequence of these two amino acids can therefore not be determined with leucine aminopeptidase. On the other hand, although the threonylaspartic acid bond is cleaved at a slower rate than the isoleucylthreonine bond, it is cleaved more rapidly than

* For a documentation of the yields in which the peptides were isolated and of the results obtained by quantitative amino acid analysis of the peptides used for the sequence studies, see the original paper.[31]

A 0-Tryp
B 0-Tryp
C 0-Tryp
D 0-Tryp
E (0-Tryp 2) Chy
F (0-Tryp 16) Chy
Chymotrypsin
Chymo trypsin
0.2M Form pH 3.1
1M Ac pH 5.1
2M Ac pH 5.1
0.2M Cit → 2M Ac
pH 3.1 → pH 5.1
2M Ac pH 6.6
0.2N Cit → 1N Cit
Cit. Ac 4.3
0.2M Form pH 3.1
0.2M Cit 1M pH 3.1
pH 3.1 → Ac pH 5.1
0.2M Form pH 3.1
1M Ac pH 5.1
2M Ac pH 5.1
[Ser.Ileu.Thr.Asp.Cys.Arg]
[Asp.Gly.Thr.Asp.Glu.Cys.Tyr]
[Glu.Ser.Tyr]
[Ser.Thr.Met]
[His.Ileu.Val.Ala.Cys.Glu.]
Gly.Asp.Pro.Tyr]
[Asp.Ala. Ser.Val]
[Val.Pro. Val.His.Phe]
alkaline hydrolysis (Leucine equivalents-mM concentration)

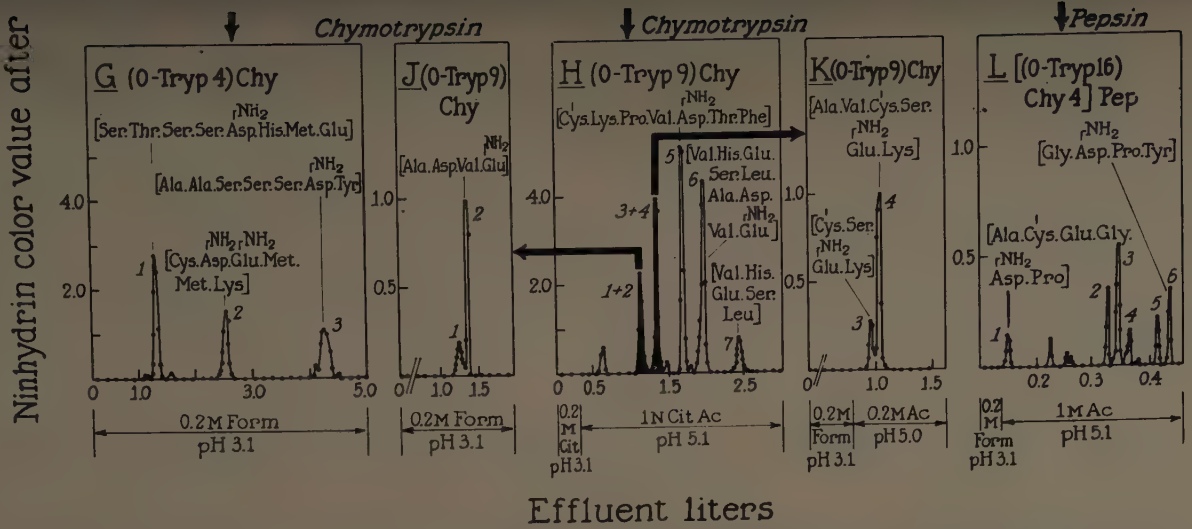


FIGURE 3. Chromatographic preparation on columns of Dowex 50-X2 of peptides from enzymatic hydrolysates of oxidized RNase.[31] The sequences of the amino acid residues within the brackets were not known at the time the peptides were isolated but have been inserted here for comparison with FIGURE 5. In the experiments shown in *3b*, *c*, *d* and *h*, the sodium form of the resin was used with sodium buffers; in all others, the ammonium form was employed with ammonium buffers. Abbreviations: Form, formic acid; Ac, acetic acid; Cit, citric acid. All columns were operated at 35° C. unless otherwise indicated. For chromatograms *b* and *l* a 610 ml. mixing chamber was used during gradient elution, while a 2460 ml. chamber was required in the remaining chromatograms. (A) A tryptic hydrolysate chromatographed on a 150 × 1.8 cm. column operated at 35° C. until 5.5 l., 50° C. thereafter. (B) A tryptic hydrolysate chromatographed on a 30 × 1.8 cm. column operated at 35° C. until 0.7 l., 50° C. thereafter. (C) Rechromatography of O-Tryp 4 plus 51 on a 150 × 1.8 cm. column operated at 25° C. (D) Rechromatography of O-Tryp 9 plus 10 on a 150 × 1.8 cm. column. (E) A chymotryptic hydrolysate of O-Tryp 2 chromatographed on a 150 × 1.8 cm. column. (F) A chymotryptic hydrolysate of O-Tryp 16 chromatographed on a 30 × 1.8 cm. column. (G) A chymotryptic hydrolysate of O-Tryp 4 chromatographed on a 150 × 1.8 cm. column. (H) A chymotryptic hydrolysate of O-Tryp 9 chromatographed on a 150 × 1.8 cm. column. (J) Rechromatography of (O-Tryp 9)Chy 1 plus 2 on a 150 × 1.8 cm. column. (K) Rechromatography of (O-Tryp 9)Chy 3 plus 4 on a 150 × 1.8 cm. column. (L) A peptic hydrolysate of (O-Tryp 16)Chy 4 chromatographed on a 150 × 0.9 cm. column. Reproduced by permission from the *Journal of Biological Chemistry*.

the aspartylcysteic acid bond. This is revealed in the initial stages of the hydrolysis by the fact that the concentration of threonine builds up less rapidly than the concentration of serine or isoleucine, but much more rapidly than the concentration of aspartic acid. Finally, the aspartylcysteic acid bond is attacked more rapidly than the peptide bond in the terminal dipeptide of cysteic acid and arginine, with the result that the concentration of free cysteic acid in the hydrolysate always was less than the concentration of free aspartic acid. The enzymatic hydrolysis therefore revealed the sequence (Ser.Ileu).-Thr.Asp.$CySO_3H$.Arg. In addition to providing this information, the experiment also revealed that aspartic acid and not asparagine was present in the peptide. This application of enzymatic hydrolysis to confirm the allocation of amide groups was of particular value in the sequence determination.

TABLE 5

PEPTIDE O-TRYP 9

NH_2 NH_2

$CySO_3H$.Lys.Pro.Val.Asp.Thr.Phe.Val.His.Glu.Ser.Leu.Ala.Asp.Val.Glu.Ala.-

NH_2

Val.$CySO_3H$.Ser.Glu.Lys

Composition
Asp, 2.04; Glu, 2.93; Ala, 2.12; Val, 3.86; Leu, 0.99; Ser, 1.98; Thr, 0.99; $CySO_3H$, 1.94; Pro, 0.99; Phe, 1.02; His, 0.99; Lys, 2.18

Impurities
None

Hydrazinolysis
Lys, 0.65

Edman degradation
First step
34%; *$CySO_3H$, 1.12*; Lys, 0.95; Pro, 1.01; Val, 3.84; Asp, 2.00; Thr, 0.97; Phe, 1.02; His, 0.95; Glu, 2.98; Ser, 1.90; Leu, 1.02; Ala, 2.14
Second step
90%; $CySO_3H$, 0.70; *Lys, 0.70*; Pro, 0.90; Val, 3.77; Asp, 1.95; Thr, 0.70; Phe, 0.99; His, 0.95; Glu, 3.25; Ser, 1.81; Leu, 0.91; Ala, 2.14
The phenylthiohydantoin of lysine was identified by paper chromatography.

The establishment of the sequence of amino acid residues in peptide O-Tryp 9 will serve to illustrate how these different procedures were applied. Reference to TABLE 5 will indicate that end-group analysis showed that peptide O-Tryp 9 contains amino-terminal cysteic acid and carboxyl-terminal lysine. In studying the results of difference analysis in this table, it will be noted that the residual peptide after the first stage of phenyl-*iso*-thiocyanate degradation had an amino acid composition that indicated that some 90 per cent of one of the two cysteic acid residues originally present had been removed as the phenylthiohydantoin. With the exception of the value for lysine, the molar ratios of the other amino acids remained unchanged. During coupling with phenyl-*iso*-thiocyanate the ϵ-amino groups of lysine in a peptide are converted to phenylthioureido groups. The partial hydrolysis of these groups with formation of free lysine during acid hydrolysis prior to amino acid analysis accounts for the recovery of some free lysine. A second stage of phenyl-*iso*-thiocyanate degra-

dation (cf. TABLE 5) demonstrated that the amino-terminal cysteic acid residue is followed by lysine.*

The further fragmentation of peptide O-Tryp 9 with chymotrypsin is illustrated in FIGURE 4, which presents the relative alignment of all the subfragments formed. The location of peptide (O-Tryp 9)Chy 5 was predicted from FIGURE 1 and confirmed by Edman degradation, which showed that this peptide, like the parent peptide O-Tryp 9, contains the amino-terminal sequence $CySO_3H$.Lys. The remaining residues in O-Tryp 9 are explainable by the sum of the residues present in peptide (O-Tryp 9)Chy 7 and (O-Tryp 9)Chy 1

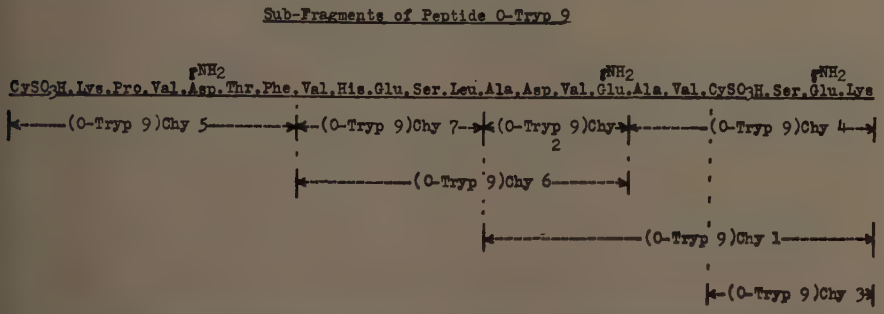


FIGURE 4. The subpeptides derived from O-Tryp 9 by chymotryptic hydrolysis (compare FIGURES 2 and 3).

TABLE 6

PEPTIDE (O-TRYP 9)CHY 5. RESIDUES 40–46

NH_2

$CySO_3H$.Lys.Pro.Val.Asp.Thr.Phe

Composition

Asp, 1.01; Val, 1.04; Thr, 0.97; $CySO_3H$, 0.91; Pro, 1.04; Phe, 1.02; Lys, 1.00; $—CONH_2$, 0.9

Impurities

None

Carboxypeptidase

Peptide conc., 0.90 mM; enzyme conc., 0.015 mg. N/ml. Amino acids released: 40 min.; Phe, 0.86; Thr, 0.29; Asp-NH_2 + peptide, 0.10; Lys, not determined; 3 major peptides: 24 hr.; Phe, 0.90; Thr, 0.79; Asp-NH_2, 0.74; Lys, 0.01; one major peptide.

Since (O-Tryp 9)Chy 4 contains lysine, this peptide represents the carboxyl end of peptide O-Tryp 9. The allocation of (O-Tryp 9)Chy 2 follows from the fact that this peptide has a sequence of residues common to both peptide (O-Tryp 9)Chy 6 and (O-Tryp 9)Chy 1.

With the relation of the subfragments of O-Tryp 9 clarified, the sequence of 22 residues in this peptide was determined by studying four peptides, (O-Tryp 9)Chy 5, (O-Tryp 9)Chy 7, (O-tryp 9)Chy 2, and (O-Tryp 9)Chy 4.

It will be seen from TABLE 6 that the action of carboxypeptidase on peptide

* Significant alterations in the molar ratios of some of the other amino acids indicates that partial acid hydrolysis had taken place during the first step. Nonetheless, the only significant component of the phenylthiohydantoin fraction was that of ε-phenylthiocarbamyllysine.

(O-Tryp 9)Chy 5 revealed the carboxyl-end sequence asparaginylthreonylphenylalanine, and that the action of the enzyme terminated with the removal of these three residues from the peptide. The failure to remove valine may have been connected with the presence of the adjacent proline residue. Edman degradation of peptide (O-Tryp 9)Chy 5 (TABLE 7) demonstrated the existence of the sequence of four amino acids, $CySO_3H$.Lys.Pro.Val. The analysis at the third stage, however, showed that both threonine and phenyl-

TABLE 7
PEPTIDE (O-TRYP 9)CHY 5. RESIDUES 40–46

$$\begin{array}{c} NH_2 \\ | \\ CySO_3H.Lys.Pro.Val.Asp.Thr.Phe \end{array}$$

Edman degradation
First step
70%; *$CySO_3H$, 0.05*; Lys, 0.45; Pro, 1.02; Val, 1.01; Asp, 1.02; Thr, 1.01; Phe, 0.99
Second step
75%; $CySO_3H$, 0.05; Lys, 0.08; Pro, 1.06; Val, 1.15; Asp, 0.92; Thr, 0.94; Phe, 1.02. The phenylthiohydantoin of phenylthiocarbamyl-lysine was detected by paper chromatography.
Third step
80%; $CySO_3H$, 0.08; Lys, not determined; *Pro, 0.15*; Val, 1.01; Asp, 0.89; Thr, 0.64; Phe, 0.70. The phenylthiohydantoin of proline was detected by paper chromatography.
Fourth step
62%; $CySO_3H$, 0.05; Lys, not determined; Pro, 0.08; *Val, 0.27*; Asp, 1.00; Thr, 0.56; Phe, 0.56

TABLE 8
PEPTIDE (O-TRYP 9)CHY 7. RESIDUES 47–51
Val.His.Glu.Ser.Leu

Composition
Glu, 1.09; Val, 0.98; Leu, 1.02; Ser, 0.93; His, 0.98; $-CONH_2$, 0.3
Impurities
Asp, 0.02; Gly, 0.05; Ala, 0.09; Lys, 0.05
Carboxypeptidase
Peptide conc., 3.5 mM; enzyme conc., 0.029 mg. N/ml. Amino acids released: 20 min. Leu, 1.00: 24 hr.; Leu, 1.00; Ser, 0.04
Leucine aminopeptidase
Peptide conc., 0.50 mM; enzyme conc., 0.150 mg. N/ml. Amino acids released: 26 hr.; Val, 0.95; Glu, 1.04; Ser, 1.01. The leucine and histidine determinations were lost.

alanine had been removed in approximately 30 per cent yield in addition to an 85 per cent removal of proline. A similar but smaller loss of threonine and phenylalanine took place in the next step. The simultaneous removal of fractional amounts of these two residues from the carboxyl end of the peptide may be attributed to partial acid hydrolysis at the asparaginylthreonine bond, an unusually labile type of peptide bond (cf. Desnuelle and Casal[39]).*

The next peptide in line, (O-Tryp 9)Chy 7, was shown by carboxypeptidase action (TABLE 8) to contain carboxyl-terminal leucine. Prolongation of the

* The threonylphenylalanine formed would be removed as the phenylthioureido derivative during extraction of the hydantoin fraction at the next step if, as is likely, cyclization of the thioureido derivative proceeded slowly.

hydrolysis with the enzyme produced a small quantity of serine, which suggested the carboxyl-terminal sequence, serylleucine. Complete hydrolysis of the peptide with leucine aminopeptidase confirmed the presence of glutamic acid, rather than glutamine. Stepwise degradation with phenyl-*iso*-thiocyanate (TABLE 9) revealed the sequence of three amino acids, Val.His.Glu: a fourth step provided evidence for the location of serine after glutamic acid, but the results showed that partial acid hydrolysis had taken place in the previous step at the serylleucine bond.

TABLE 9
PEPTIDE (O-TRYP 9)CHY 7. RESIDUES 47–51
Val.His.Glu.Ser.Leu

Edman degradation
First step
90%; *Val. 0.09*; His, 0.83; Glu, 1.11; Ser, 1.03; Leu, 1.03
Second step
89%; Val, 0.02; *His, 0.23*; Glu, 1.06; Ser, 1.00; Leu, 0.95
Third step
36%; Val, <0.01; His, 0.25; *Glu, 0.41*; Ser, 0.99; Leu, 1.01
Fourth step
49%; Val, <0.01; His, 0.25; Glu, 0.31; *Ser, 0.40*; Leu, 0.58

TABLE 10
PEPTIDE (O-TRYP 9)CHY 2. RESIDUES 52–55

$$\begin{array}{c} NH_2 \\ | \\ \text{Ala.Asp.Val.Glu} \end{array}$$

Composition
Asp, 0.76; Glu, 0.89; Ala, 1.07; Val, 1.04; $-CONH_2$, 1.6
Impurities
Gly, 0.08; Ser, 0.18; Thr, 0.01; Lys, 0.15
Leucine aminopeptidase
Peptide conc., 1.0 mM; enzyme conc., 0.015 mg. N/ml. Amino acids released: 24 hr.; Ala, 1.09; Asp, 0.93; Val, 1.09; Glu-NH_2, 0.89
Edman degradation
First step
85%; *Ala, 0.18*; Asp, 0.95; Val, 1.06; Glu, 0.99
Second step
43%; Ala, 0.08; *Asp, 0.08*; Val, 1.01; Glu, 0.99
Third step
40%; Ala, 0.09; Asp, 0.09; *Val, 0.31*; Glu, 1.00

Following peptide (O-Tryp 9)Chy 4 is (O-Tryp 9)Chy 2, a tetrapeptide, evidence for the sequence of which is provided in TABLE 10. Upon leucine aminopeptidase hydrolysis (O-Tryp 9)Chy 2 yielded glutamine, confirming the presence of the single amide group shown by the analysis of the peptide. The sequence of the four amino acids was established by the results of the Edman degradation.

This leaves the remaining subpeptide of O-Tryp 9, (O-Tryp 9)Chy 4. As inspection of TABLE 11 will show the action of leucine aminopeptidase on this peptide produced glutamine, confirming the presence of the amide group de-

tected by analysis of the peptide. Overlapping peptides at the valine position on effluent curves from the amino acid analyzer complicated the results of the analysis of the rate of release of free amino acids by leucine aminopeptidase. The results nevertheless furnished evidence to show that glutamine is preceded by serine and cysteic acid. Information obtained by stepwise degradation with phenyl-*iso*-thiocyanate (TABLE 12) demonstrated the sequence of four amino acids, Ala.Val.$CySO_3H$.Ser. The presence of carboxyl-terminal lysine was already known from the results obtained by end-group analysis of O-Tryp

TABLE 11

PEPTIDE (O-TRYP 9)CHY 4. RESIDUES 56–61

$$\begin{array}{c} NH_2 \\ | \\ \text{Ala.Val.CySO}_3\text{H.Ser.Glu.Lys} \end{array}$$

Composition
Glu, 0.94; Ala, 0.96; Val, 1.03; Ser, 0.79; $CySO_3H$, 1.00; Lys, 1.06; $-CONH_2$, 1.0
Impurities
None
Leucine aminopeptidase
Peptide conc., 0.85 mM; enzyme conc., 0.015 mg. N/ml. Amino acids released: 25 hr.; Ala, 0.60; $CySO_3H$. 0.13; Ser, 0.13; Glu-NH_2, 0.08. The free valine present was overlapped by a peptide, and there was also a peptide at the lysine position on the chromatograms. Two additional peptides were noted.

TABLE 12

PEPTIDE (O-TRYP 9)CHY 4. RESIDUES 56–61

$$\begin{array}{c} NH_2 \\ | \\ \text{Ala.Val.CySO}_3\text{H.Ser.Glu.Lys} \end{array}$$

Edman degradation
First step
25%; *Ala, <0.01*; Val, 1.37; $CySO_3H$, determination lost; Ser, 1.06; Glu, 0.97; Lys, not determined
Second step
55%; Ala, <0.01; *Val, 0.10*; $CySO_3H$, 1.03; Ser, 1.00; Glu, 0.97; Lys, not determined
Third step
57%; Ala, <0.01; Val, <0.01; *$CySO_3H$, 0.21*; Ser, 1.00; Glu, 1.00; Lys, not determined
Fourth step
80%; Ala, < 0.01; Val, <0.01; $CySO_3H$, 0.17; *Ser, 0.55*; Glu, 1.00; Lys, not determined

9, and the results of both leucine aminopeptidase hydrolysis and stepwise degradation thus define the final dipeptide sequence as glutaminyllysine.

It must suffice to say that the amino acid sequences of the remaining peptides isolated for detailed analysis were investigated[38] by similar procedures, and once established, were fitted together on the basis of the results summarized in FIGURES 1 and 2 to give, ultimately, the total sequence shown in FIGURE 5. If it is assumed that normal peptide bonds (or at least bonds capable of isomerization to normal peptide bonds),[1] link all the residues and if, further, the protein is assembled in its entirety by the condensation of L-amino acids, the sequence shown in FIGURE 5 automatically defines the structural formula of performic acid-oxidized RNase (this would include oxazoline structures of the type proposed by Rydon[40] and ester bonds at serine and threonine residues).

The Disulfide Bonds

The final step in reconstructing the covalent structure of the native enzyme molecule is the establishment of the location of the disulfide bonds. An attack on this problem demands in the first place a procedure for degrading RNase into peptides under conditions that preclude disulfide bond interchange.[41] The resistance of native RNase to the action of trypsin and chymotrypsin led Spackman[42] in his initial experiments to use these enzymes in solutions 2 *M* in guani-

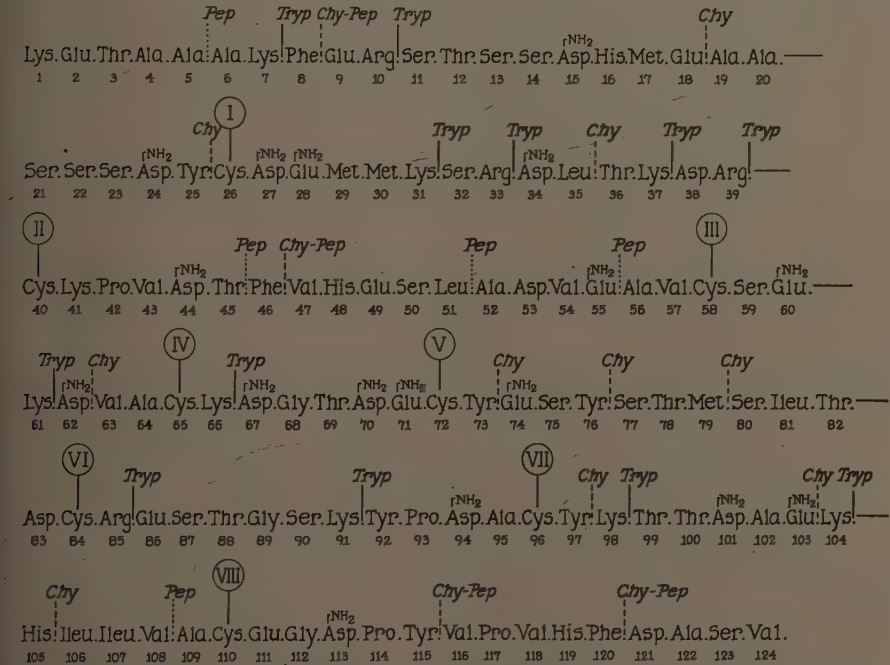


FIGURE 5. The sequence of amino acid residues in oxidized RNase.[38] The single chain of 124 residues is broken into segments that should be imagined to be joined end to end. Methionine sulfone and cysteic acid residues are represented by Met and Cys respectively. Following the system described in FIGURE 1, the bonds at which the protein is attacked by proteolytic enzymes are designated by vertical lines. Reproduced by permission from the *Journal of Biological Chemistry*.

dinium chloride and containing *N*-ethyl-maleimide as an agent for the trapping of traces of free thiols, the presence of which catalyzes the disulfide interchange reaction at neutral or higher *p*H values. The guanidinium chloride was subsequently removed in a desalting step, following which the peptides were fractionated on columns of Dowex 50-X2. Evidence for only two of the possible four disulfide bonds in the form of cystine peptides was obtained. The failure to detect appreciable quantities of peptides corresponding to the remaining two disulfide bonds led to an investigation of the factors influencing the disulfide interchange reaction in the model system cystine-glutathione. Spackman found[43] that the presence of *N*-ethyl-maleimide failed to prevent the formation of the mixed disulfide of cystine and glutathione, particularly at *p*H values

over 7, and over-all loss of disulfide occurred in the presence of stoichiometric amounts of *N*-ethyl-maleimide probably because the equilibrium

$$R—S—S—R + OH^- \rightleftharpoons R—S^- + R—S—OH$$

was being pulled to the right. The presence of 2 *M* guanidinium chloride promoted the occurrence of oxidative side reactions that led to the destruction of disulfide. Disulfide stability increased notably at low *p*H values.

This experience led to the adoption of an approach involving a preliminary peptic hydrolysis of ribonuclease at *p*H 2.[43] After approximately 15 peptide bonds had been broken, the *p*H was increased to near 7 and the degraded protein subjected to further proteolysis successively with trypsin and chymotrypsin. The peptides formed were fractionated on columns of Dowex 50-X2, and the zones in the effluent containing cystine-positive material were collected. Following desalting by ion exchange, the cystine-positive fractions were oxidized with performic acid, and the reaction products from each fraction were again chromatographed on columns of Dowex 50-X2. Each cystine-containing peptide gave rise to two cysteic acid-containing peptides: the introduction of a sulfonic acid group into each of the half-members of the original peptide resulted in a decrease in affinity for Dowex 50, while the unaffected, cystine-free components of the mixture of peptides present in the cystine-positive zones chromatographed with much greater retention under the same conditions of elution. Quantitative amino acid analyses of the cysteic acid peptides showed that the cystine-containing peptides had been formed in yields of 20 to 50 per cent from the protein. The analyses of the cysteic acid peptides were also by themselves sufficient, without the need for sequence determination, when considered in conjunction with the information from FIGURE 5 to clearly define the location of the disulfide bonds in RNase. These bonds connect the pairs of half-cystine residues I–VI, II–VII, III–VIII, and IV–V. The result shown in FIGURE 6, which represents an abbreviated structural formula for RNA, was obtained by bringing together the appropriate half-cystine residues in FIGURE 5 and folding the chain around into a two-dimensional coil.

Secondary and Tertiary Structure

The X-ray structure analysis of crystals of RNase has progressed steadily, but detailed information of the kind available for myoglobin[44] and hemoglobin[45] is not likely to be derived for some time. Investigations of the secondary and tertiary structure of RNase have therefore had to rely on less discriminating procedures. Much of what is known has been derived from measurements of properties that represent the average of the separate contributions of many structural elements and therefore lacks the resolution that would be most helpful for probing the secondary and tertiary structure. Indeed it is evident that considerable future effort will have to be directed at isolating the contributions of many of the individual functional elements defined in FIGURE 6 before an integration of their separate contributions into the properties of the whole becomes feasible. In spite of this, it is of interest to examine some of the implications of FIGURE 6 in the light of information already available about the secondary and tertiary structure.

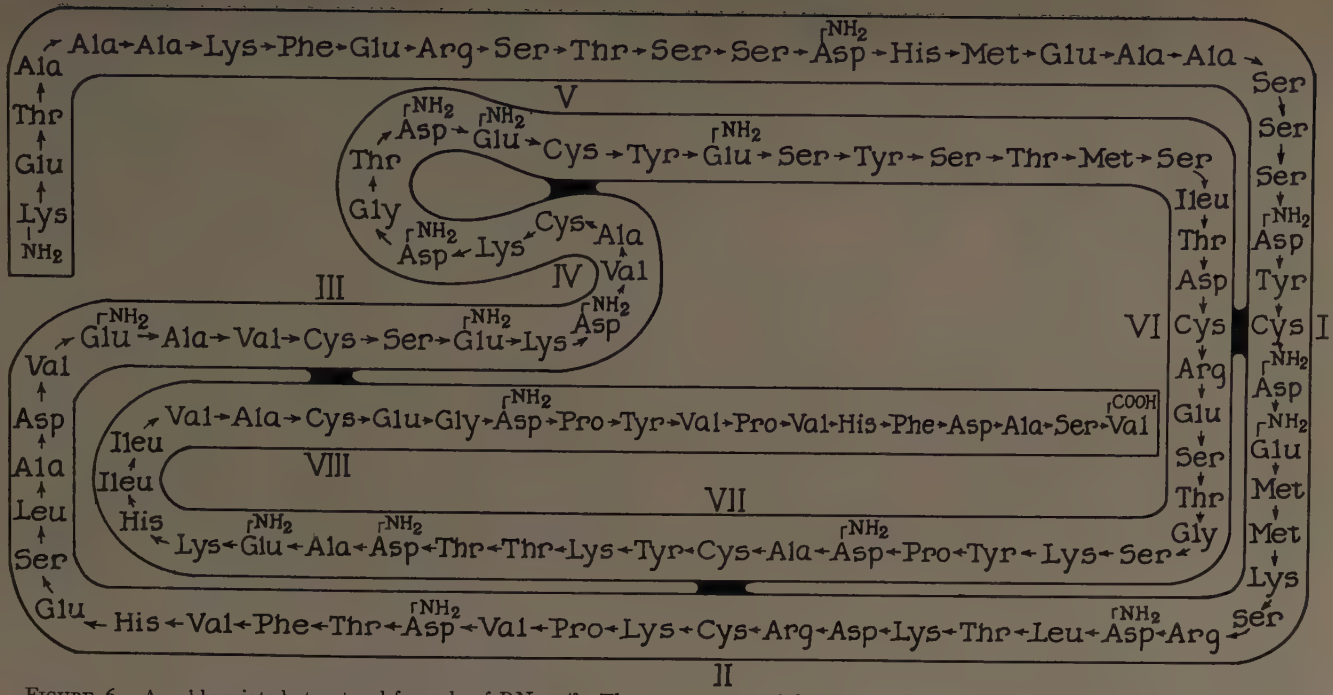


FIGURE 6. An abbreviated structural formula of RNase.[43] The arrangement of the peptide chain in a two-dimensional coil is for purposes of presentation and does not indicate knowledge of aspects of the secondary and tertiary structure. Reproduced by permission from the *Journal of Biological Chemistry*.

It is at this point that the use of atomic models is of value in examining the possible conformations that may be assumed by a molecule of RNase. An interesting, though admittedly extremely preliminary, speculation along these lines is shown in FIGURES 7 and 8, which present photographs taken from different aspects of a model of RNase built in Scheraga's laboratory[46] from accurately fabricated components embodying the bond distances and bond angles

FIGURE 7. Model of RNase according to Scheraga[46] with possible secondary and tertiary structure shown. Hydrogen atoms have been omitted. Reproduced by permission from the *Journal of The American Chemical Society.*

FIGURE 8. View from the left of the model shown in FIGURE 7. Reproduced by permission from the *Journal of The American Chemical Society*.

worked out by Pauling and Corey (cf. Lindley and Rollett[47]) from X-ray diffraction analyses of the crystal structure of simple amino acids and peptides.

The main features to be noted in FIGURES 7 and 8 are the carboxyl-terminal end of the chain (in the right upper portion of FIGURE 7), the amino-terminal segment of 20 residues (emphasized with the vertical white cylinder visible in both photographs), and the loop introduced by the IV-V disulfide bond (the cluster of residues at the far right center of FIGURE 7). The disulfide bonds are not all visible in the photographs, but the II–VII disulfide bond is closest to the camera in FIGURE 8 slightly below the center of the picture. The peptide chain may be traced by following the direction of the helical elements that have been emphasized with the white cylinders.

A model of this kind should seek to explain a considerable specificity in the side-chain interactions. The important observations of White and Anfinsen[48] with mercaptoethanol-reduced RNase have revealed that such a specificity must exist. The fully reduced protein is transformed in high yield upon gentle oxidation to a material that is indistinguishable from the parent enzyme. Stabilization of the secondary and tertiary structure is therefore not due to the presence of disulfide bonds alone.

Richards[49] has observed that limited proteolysis with subtilisin converts RNA to RNase S, a protein which appears to differ in primary structure from RNase solely in the absence of a peptide bond between residues 20 and 21. The amino-terminal peptide fragment of 20 residues interacts highly specifically with the protein core of the molecule, as indicated by the observation that RNase S is essentially as active enzymatically as RNase, while the S peptide and the protein core, which have an exceptionally high but reversible association constant, are individually inactive. The model attributes this strong and specific interaction to the formation of 7 hydrogen and 2 hydrophobic bonds between the helical form of peptide S and the rest of the molecule.

Peptic hydrolysis of RNase at *p*H 1.7 has been observed by Anfinsen[50] (cf. also Ginsberg and Schachman[51]) to cause complete inactivation of the enzyme when proteolysis is restricted, apparently, to a single peptide bond between residues 120 and 121. Gundlach *et al.*[52] have demonstrated that carboxymethylation of RNase at *p*H 5 with iodoacetic acid can lead to the formation of an inactive mono-carboxymethyl RNase in which substitution is at one of the four histidine residues. On the other hand, Barnard and Stein[53] have indicated that reaction of RNase with bromoacetate at *p*H 7 can produce a mono-carboxymethyl RNase and that the reaction is inhibited by the presence of cytidylate. These workers provided evidence to demonstrate that their inactive derivative was carboxymethylated at the histidine residue at position 119.

These and other observations (cf. Anfinsen[54]) on the inactivation of RNase imply that structural elements associated with either the binding of substrate or of catalysis, or both, are at least in part contributed by the carboxyl end of the peptide chain. The inactivation associated with the removal of S peptide from RNase S implies that peptide S may also contain elements of the binding and catalytic sites or that attachment of S peptide serves to maintain the configuration of these sites. The Scheraga model accommodates these ob-

servations by bringing the amino and carboxyl ends of the peptide chain into apposition but, as yet, cannot hope to reveal the function of peptide S.

The model has dimensions that agree well with the dimensions of the unit cell in crystals of RNase[55] as well as with deductions reached from measurements of the hydrodynamic parameter β of Scheraga.[56] The highly compact structure suggested by physical measurements is in accord with observations that reveal the existence of shielding and hinderance in the molecule of the native protein.

Titration studies have amply demonstrated that three of the six tyrosine residues in RNase titrate abnormally (see Cha and Scheraga for summary[57]). The average pK′ for the 11 carboxyl groups is approximately 0.5 pK units less than the pK′ of acetic acid. When the protein is dissolved in strong solutions of urea or guanidinium chloride, all six of the phenolic hydroxyl groups of tyrosine in the protein titrate with a pK′ very close to that observed for phenol in the same solvents, while the average pK′ of the carboxyl groups increases to a value very close to that observed for acetic acid in these solvents.

Optical rotatory dispersion measurements of aqueous solutions of RNase[58] indicate that the addition of urea to such solutions brings about a large increase in negative rotation, an effect consistent with the view that a transition from a relatively more helical to a more random coil configuration takes place under the influence of the hydrogen bond-breaking action of urea.

A relatively abrupt increase in negative rotation is observed when RNase solutions are heated above a temperature of 60° C.[56] The abrupt transition is reversed on cooling. The passage of the molecule from a more folded to a less compact configuration with heating would probably be reflected by this kind of a change in optical rotation. The existence of a transition above 60° C. is also revealed by the observation that heating about the transition temperature is necessary to bring about complete replacement of all the exchangeable hydrogen atoms in RNase with deuterium.[59]

Deuterium exchange studies (compare Schildkraut and Scheraga for summary[59]) have shown that all the exchangeable hydrogen atoms in performic acid-oxidized RNase exchange instantly with solvent deuterium. The native enzyme, however, contains some 70 hydrogen atoms (out of a total of 245 exchangeable hydrogen atoms) that either fail to exchange at all or exchange slowly.

Comparison of the nuclear magnetic resonance spectrum of RNase in deuterium oxide solution in the presence and absence of urea[60] indicates that there is an enhancement in the freedom in the internal motions of the molecule in the presence of urea. This enhancement is particularly evident in the nuclear magnetic resonance spectrum of oxidized RNase in deuterium oxide.

The viscosity of solutions of RNase[56] increases to a value comparable to that found for solutions of oxidized RNase in water when urea is added.

To explain the high degree of compactness in the structure to which these observations attest, Scheraga relied on the α-helix of Pauling *et al.*[61] for major portions of the model. The degree of helicity to be ascribed was decided by making allowance for nonhydrogen-bonded structural elements, for example, the loop involving the IV–V disulfide bond, the portions of the structure-round

proline residues, and the major folding requirements set by FIGURE 6, and by assuming that, on average,[62] there are 4 hydrogen bonds at the ends of each helix capable of rapid exchange with solvent hydrogen. On this basis, the model shown in FIGURES 7 and 8 has 30 peptide bond-hydrogen atoms in non-hydrogen-bonded structures and 24 exchangeable hydrogen atoms at the ends of helices, a total of 54 readily exchangeable peptide hydrogen atoms. Since there are 124 residues in the peptide chain, approximately 70 peptide bond-hydrogen atoms in hydrogen-bonded form remain for a relatively slow exchange with solvent hydrogen.

Although admittedly highly conjectural, a model of the kind proposed by Scheraga has the virtue of making an attempt to bring together a considerable body of information and serves as a guide for more refined studies in the future that we may hope will greatly narrow the limits of uncertainty in our present knowledge of the secondary and tertiary structure of the protein.

Acknowledgments

My participation in the investigation of the primary structure of RNase was made possible by the guidance and encouragement of Stanford Moore and William H. Stein during an association of almost one decade in their laboratory at The Rockefeller Institute, New York, N.Y. Their unfailing support is gratefully acknowledged.

The concluding steps in the determination of the amino acid sequence of RNase were accomplished at Brookhaven National Laboratory, Upton, N.Y., under the auspices of the United States Atomic Energy Commission.

I am grateful to M. A. Mitz of Armour and Company, Chicago, Ill., for maintaining the supply of the particular preparation of crystalline RNase used in the course of the amino acid sequence studies. Approximately 8.5 gm. of the protein were consumed in this determination.

In a project involving as many quantitative amino acid analyses as the studies reported, a reasonable rate of progress could be maintained only by automation of the analytical method. It was largely in anticipation of the analytical demands of the RNase problem that the apparatus described by Spackman *et al.*[43] was developed. I am greatly indebted to D. H. Spackman for his advice on construction of an amino acid analyzer.

I wish to thank H. A. Scheraga for permission to present details of the RNase model prior to publication.

References

1. SANGER, F. & H. TUPPY. 1951. Biochem. J. **49:** 463, 481.
2. SANGER, F. & E. O. P. THOMPSON. 1953. Biochem. J. **53:** 353, 366.
3. SHEPHERD, R. G., S. D. WILSON, K. S. HOWARD, P. H. BELL, D. S. DAVIES, S. B. DAVIS, E. A. EIGNER & N. SHAKESPEARE. 1956. J. Am. Chem. Soc. **78:** 5067.
4. LI, C. H. 1956. Advances in Protein Chemistry. **11:** 101.
5. WHITE, W. F. & W. A. LANDMANN. 1955. J. Am. Chem. Soc. **77:** 1711.
6. BROMER, W. W., L. G. SINN & O. K. BEHRENS. 1957. J. Am. Chem. Soc. **79:** 2807.
7. GESCHWIND, I. I., C. H. LI & L. BARNAFI. 1957. J. Am. Chem. Soc. **79:** 620, 6394.
8. HARRIS, J. I. & P. ROOS. 1959. Biochem. J. **71:** 434, 445.
9. HARRIS, J. I. 1959. Biochem. J. **71:** 451. Nature. **184:** 167.
10. JONES, W. 1920. Am. J. Physiol. **52:** 203.
11. NORTHROP, J. H., M. KUNITZ & R. M. HERRIOTT. 1948. *In* Crystalline Enzymes. 2nd ed. Columbia Univ. Press. New York, N.Y.

12. Rothen, A. 1940. J. Gen. Physiol. **24:** 203.
13. Martin, A. J. P. & R. R. Porter. 1951. Biochem. J. **49:** 215.
14. Hirs, C. H. W., S. Moore & W. H. Stein. 1953. J. Biol. Chem. **200:** 493.
15. Schmidt, G., R. Cubiles, N. Zöllner, L. Hecht, N. Strickler, M. Seraidarian & S. J. Thannhauser. 1951. J. Biol. Chem. **192:** 715.
16. Markham, R. & J. D. Smith. 1952. Biochem. J. **52:** 552.
17. Volkin, E. & W. E. Cohn. 1953. J. Biol. Chem. **205:** 767.
18. Brown, D. M. & A. R. Todd. 1953. J. Chem. Soc. : 2040.
19. Heppel, L. A. & P. R. Whitfeld. 1955. Biochem. J. **60:** 1.
20. Kunitz, M. 1953. Biochem. Preparations. **3:** 9.
21. Raacke, I. D. & C. H. Li. 1954. Biochim. et Biophys. Acta. **14:** 290.
22. King, T. P. & L. C. Craig. 1958. J. Am. Chem. Soc. **80:** 3366.
23. Hirs, C. H. W., W. H. Stein & S. Moore. 1954. J. Biol. Chem. **211:** 941.
24. Anfinsen, C. B., R. R. Redfield, W. L. Choate, J. Page & W. R. Carroll. 1954. J. Biol. Chem. **207:** 201.
25. Hirs, C. H. W. 1956. J. Biol. Chem. **219:** 611.
26. Hirs, C. H. W., S. Moore & W. H. Stein. 1956. J. Biol. Chem. **219:** 623.
27. Moore, S. & W. H. Stein. 1954. J. Biol. Chem. **211:** 893.
28. Spackman, D. H., W. H. Stein & S. Moore. 1958. Anal. Chem. **30:** 1190.
29. Hirs, C. H. W., W. H. Stein & S. Moore. 1956. J. Biol. Chem. **221:** 151.
30. Bailey, J. L., S. Moore & W. H. Stein. 1956. J. Biol. Chem. **221:** 143.
31. Hirs, C. H. W. 1960. J. Biol. Chem. **235:** 625.
32. Redfield, R. R. & C. B. Anfinsen. 1956. J. Biol. Chem. **221:** 393.
33. Klee, W. A. & F. M. Richards. 1957. J. Biol. Chem. **229:** 489.
34. Edman, P. 1950. Acta Chem. Scand. **4:** 283. 1953. **7:** 700.
35. Braunitzer, G. 1955. Chem. Ber. **88:** 2025.
36. Green, N. M. & H. Neurath. 1954. *In* The Proteins. II. H. Neurath & K. Bailey, Eds. Academic Press. New York, N.Y.
37. Smith, E. L. & D. H. Spackman. 1955. J. Biol. Chem. **212:** 271.
38. Hirs, C. H. W., S. Moore & W. H. Stein. 1960. J. Biol. Chem. **235:** 633.
39. Desnuelle, P. & A. Casal. 1948. Biochim. et Biophys. Acta. **2:** 64.
40. Rydon, H. N. 1958. Nature. **182:** 928.
41. Ryle, A. P. & F. Sanger. 1955. Biochem. J. **60:** 535.
42. Hirs, C. H. W., W. H. Stein & S. Moore. 1958. *In* Symposium on Protein Structure. A. Neuberger, Ed. Wiley. New York, N.Y.
43. Spackman, D. H., W. H. Stein & S. Moore. 1960. J. Biol. Chem. **235:** 648.
44. Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips & V. C. Shore. 1960. Nature. **185:** 422.
45. Perutz, M. F., M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will & A. C. T. North. 1960. Nature. **185:** 416.
46. Scheraga, H. A. J. Am. Chem. Soc. In press.
47. Lindley, H. & J. S. Rollett. 1955. Biochim. et Biophys. Acta. **18:** 183.
48. White, F. H. 1960. Federation Proc. **19:** 333.
49. Richards, F. M. & P. J. Vithayathil. 1959. J. Biol. Chem. **234:** 213.
50. Anfinsen, C. B. 1956. J. Biol. Chem. **221:** 405.
51. Ginsburg, A. & H. K. Schachman. 1960. J. Biol. Chem. **235:** 115.
52. Gundlach, H. G., W. H. Stein & S. Moore. 1959. J. Biol. Chem. **234:** 1754.
53. Barnard, E. A. & W. D. Stein. 1959. Biochem. J. **71:** 14P.
54. Anfinsen, C. B. 1959. *In* The Molecular Basis of Evolution. Wiley. New York, N.Y.
55. Carlisle, C. H. & H. Scouloudi. 1951. Proc. Roy. Soc. London. **A207:** 496.
56. Harrington, W. F. & J. A. Schellman. 1956. Compt. rend. trav. Lab. Carlsberg. Ser. Chim. **30:** 21.
57. Cha, C-Y. & H. A. Scheraga. 1960. J. Am. Chem. Soc. **82:** 54.
58. Sela, M., C. B. Anfinsen & W. F. Harrington. 1957. Biochim. et Biophys. Acta. **26:** 502.
59. Schildkraut, C. A. & H. A. Scheraga. 1960. J. Am. Chem. Soc. **82:** 58.
60. Kowalsky, A. & P. D. Boyer. 1960. Federation Proc. **19:** 337.
61. Pauling, L., R. B. Corey & H. R. Branson. 1951. Proc. Natl. Acad. Sci. U. S. **37:** 205.
62. Schellman, J. A. 1955. Compt. rend. trav. Lab. Carlsberg. Ser. Chim. **29:** 230.

SPECIFIC RECOMBINATION OF THE SUBUNITS OF HEMOGLOBIN

Harvey A. Itano and Elizabeth Robinson

The National Institute of Arthritis and Metabolic Diseases of the National Institutes of Health, Public Health Service, Bethesda, Md.

The subunits of hemoglobin, its four polypeptide chains, are exchanged between molecules when a mixture of two or more hemoglobins is dissociated and recombined. While it is true that other proteins are composed of subunits, certain factors favor use of human hemoglobin in experiments on recombination. In general, each hemoglobin is composed of two types of polypeptide chains, and inherited variants with chemical defects in one or the other chain are available. Moreover, preparation of specifically labeled molecules is relatively simple because one heme is bound to each chain and because immature red cells readily incorporate labeled amino acids into hemoglobin. The physical property of hemoglobin upon which practicability of recombination experiments rests is its reversible dissociation under mild conditions.

The apparent genetic paradox that gave rise to these studies and the genetic implications of our results have been summarized.[1] We shall review here the evidence for exchange of subunits and for mode of exchange, and then consider ways in which the method of recombination can be applied to chemical studies of hemoglobin.

Methods

Procedures for preparing samples, specific labeling, and recombining mixtures have been described elsewhere.[2,3,4] A sample is dissociated by exposure for varying lengths of time to high or low *p*H and recombined by addition of a neutralizing buffer or by dialysis against a buffer of neutral *p*H. The final composition of a mixture thus recombined is compared with its composition before treatment. Each sample is also dissociated and reassociated in the absence of other components. Electrophoresis by the moving boundary and starch-block methods and column chromatography have been used to determine these compositions. Changes in composition not apparent by these methods alone may be detected by separating labeled components by starch-block electrophoresis or chromatography and then measuring the radioactivity of the individual components.[3,4]

Nomenclature

Carbonmonoxyhemoglobin (HbCO), the addition compound of carbon monoxide to ferrohemoglobin, was the compound used in most of our experiments. One positive charge was added to each heme at neutral or acidic *p*H by oxidation of ferrohemoglobin to ferrihemoglobin (Hb^+). This charge, which alters the electrophoretic mobility of the molecule,[5] was used as a specific heme label in some of the experiments. The two types of chains of human hemoglobin are identified by their *N*-terminal amino acid sequences, val-leu- for the α-chain and val-his-leu- for the β-chain; and a molecule is composed of a pair of

each of these chains.[6] The α-chain and the β-chain carry the chemical abnormalities of hemoglobins I and S, respectively; and the abnormality of hemoglobin C occurs in the same position of the β-chain as that of hemoglobin *S*.[7] Examples of nomenclature in terms of chains and charge are shown in TABLES 1 and 2, the superscript A denoting a normal chain of adult hemoglobin and any other superscript denoting the defective chain of the corresponding abnormal hemoglobin. The heme label of ferrihemoglobin is indicated by a plus sign after the superscript, and increments in net charge of subunits and molecules are placed outside brackets. The latter figures, which represent increments in charge with respect to those of the CO form of the subunits and molecule of hemoglobin A, are one per chain for Hb^+ and $+1$, $+2$, and -2 per chain, respectively, for the amino acid substitutions of the defective chains of

TABLE 1

SYMBOLS FOR SUBUNITS AND MOLECULES OF CARBONMONOXYHEMOGLOBINS (HbCO)*

Hb	α-chain	β-chain	HbCO
A	α^{A}	β^{A}	$\alpha_2{}^{A}\beta_2{}^{A}$
S	α^{A}	$(\beta^{S})^{+1}$	$(\alpha_2{}^{A}\beta_2{}^{S})^{+2}$
I	$(\alpha^{I})^{-2}$	β^{A}	$(\alpha_2{}^{I}\beta_2{}^{A})^{-4}$

* For comparison of these symbols with those used in earlier papers see Itano *et al.*[2,3,10]

TABLE 2

SYMBOLS FOR SUBUNITS AND MOLECULES OF FERRIHEMOGLOBINS (Hb^+)

Hb	α^+-chain	β^+-chain	Hb^+
A	α^{A+}	β^{A+}	$(\alpha_2^{A+}\beta_2^{A+})^{+4}$
S	α^{A+}	$(\beta^{S+})^{+2}$	$(\alpha_2^{A+}\beta_2^{S+})^{+6}$
C	α^{A+}	$(\beta^{C+})^{+3}$	$(\alpha_2^{A+}\beta_2^{C+})^{+8}$

hemoglobins S, C, and I. In accordance with results of sedimentation studies, dissociation both in acid and alkali is assumed to result in particles of one half the mass of hemoglobin.[8,9] Dissociation of $\alpha_2\beta_2$ into two identical half molecules $\alpha\beta$ is termed symmetrical dissociation, and dissociation into unlike chain-pairs α_2 and β_2 is termed asymmetrical dissociation (FIGURE 1).

Results

Recombination of acid-dissociated HbCO A, S, and C. No change in composition was observed when mixtures of these hemoglobins were recombined, although each dissociates reversibly in acid.[2] Therefore, if dissociation is symmetrical, there is no evidence that unlike half-molecules $\alpha^{A}\beta^{A}$, $(\alpha^{A}\beta^{S})^{+1}$, and $(\alpha^{A}\beta^{C})^{+2}$ combine with each other to form hybrid molecules $(\alpha_2^{A}\beta^{A}\beta^{S})^{+1}$, $(\alpha_2^{A}\beta^{A}\beta^{C})^{+2}$, and $(\alpha_2^{A}\beta^{S}\beta^{C})^{+3}$. If dissociation is asymmetrical, exchange of the common chain-pair α_2 among the molecules $\alpha_2^{A}\beta_2^{A}$, $(\alpha_2^{A}\beta_2^{S})^{+2}$, and $(\alpha_2^{A}\beta_2^{C})^{+4}$ would not be detectable. Finally, if dissociated mixtures contain single chains,

absence of change in composition indicates that unlike β-chains do not pair with each other on recombination.

Recombination of acid-dissociated HbCO and Hb$^+$ (see Singer *et al.*)[8]. A single new component of intermediate mobility is produced when HbCO and Hb$^+$ prepared from the same hemoglobin are recombined. The same result was obtained whether hemoglobin A, S, or C was used. Patterns obtained with hemoglobin

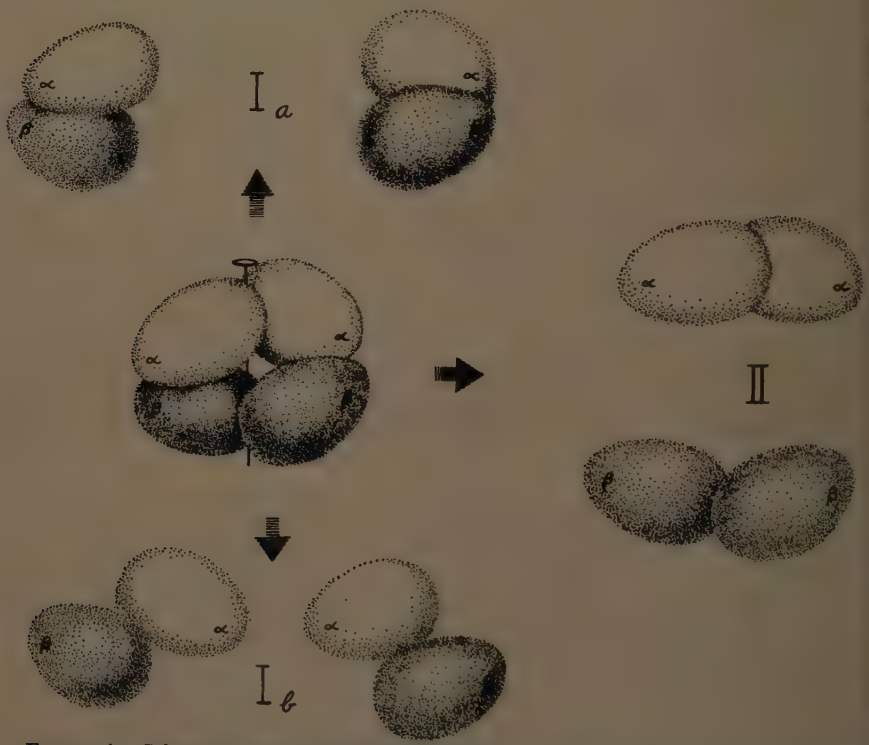


FIGURE 1. Schematic model of a molecule of hemoglobin showing the two possible modes (I*a* and I*b*) of dissociation into $\alpha\beta + \alpha\beta$ and dissociation into chain-pairs $\alpha_2 + \beta_2$ (II). In order to show the dyad symmetry of the molecule, the α-chains and β-chains have been separated slightly along the dyad.

C are shown in FIGURE 2*a*. Either symmetrical or asymmetrical recombination can account for the appearance of one intermediate component. Thus,

$$(\alpha_2^A\beta_2^C)^{+4} + (\alpha_2^{A+}\beta_2^{C+})^{+8} = 2(\alpha^A\alpha^{A+}\beta^C\beta^{C+})^{+6}, \text{ symmetrical}$$

or

$$(\alpha_2^A\beta_2^C)^{+4} + (\alpha_2^{A+}\beta_2^{C+})^{+8} = (\alpha_2^A\beta_2^{C+})^{+6} + (\alpha_2^{A+}\beta_2^C)^{+6}, \text{ asymmetrical.}$$

The significance of this experiment is that species such as $(\alpha^A\alpha^{A+}\beta_2^C)^{+5}$ and $(\alpha^A\alpha^{A+}\beta_2^{C+})^{7}$ are not observed. Molecules in which one to three of the hemes are in the ferric state exist as intermediate compounds in the oxidation of HbCO to Hb$^+$;[5] therefore their absence in recombined mixtures indicates that

in all probability dissociation into single chains does not occur in acid. Their absence shows also that exchange of electrons and CO molecules between subunits or molecules does not occur in these experiments. Although the sedimentation behavior of Hb^+ was not investigated, it is evident from formation of

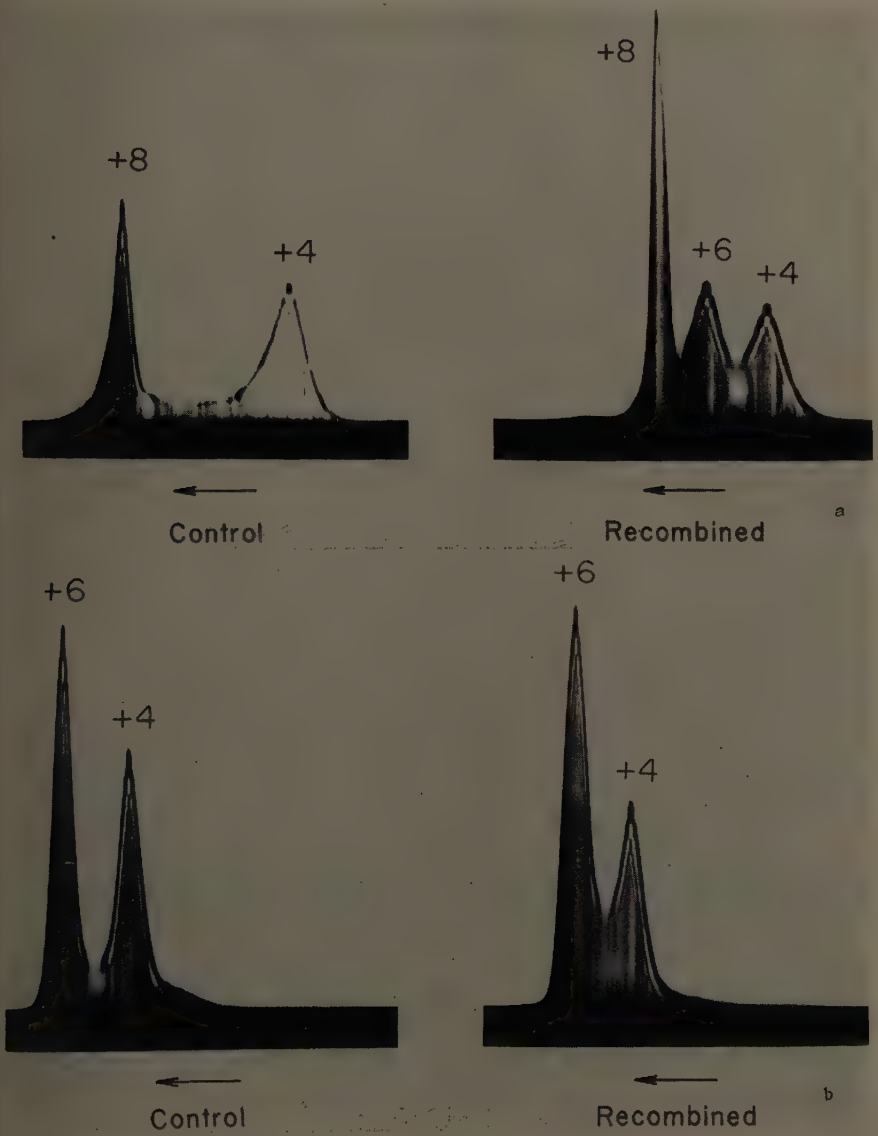


FIGURE 2*a*. Electrophoretic patterns of control and recombined mixtures of HbCO C and Hb^+ C. From Singer and Itano.[3]

FIGURE 2*b*. Electrophoretic patterns of control and recombined mixtures of HbCO C and Hb^+ S.

the intermediate component that the Hb^+ compound of hemoglobins A, S, or C, like the HbCO compound, dissociates in acid.

No new electrophoretic component is observed when Hb^+ A is recombined with HbCO S or when Hb^+ S is recombined with HbCO C. These results are consistent with asymmetrical recombination,

$$(\alpha_2^{A+}\beta_2^{A+})^{+4} + (\alpha_2^{A}\beta_2^{S})^{+2} = (\alpha_2^{A+}\beta_2^{S})^{+4} + (\alpha_2^{A}\beta_2^{A+})^{+2},$$

and

$$(\alpha_2^{A+}\beta_2^{S+})^{+6} + (\alpha_2^{A}\beta_2^{C})^{+4} = (\alpha_2^{A+}\beta_2^{C})^{+6} + (\alpha_2^{A}\beta_2^{S+})^{+4}.$$

The latter experiment is shown in FIGURE 2*b*. Absence of an intermediate component $(\alpha^A\alpha^{A+}\beta^C\beta^{S+})^{+5}$, rules out symmetrical recombination. Exchange of equivalent chain-pairs $(\beta_2^{S+})^{+4}$ and $(\beta_2^{C})^{+4}$ without formation of a new electrophoretic component is analogous to exchange of the chain-pair α_2^A when HbCO A, S, and C are recombined; however, it would be possible in the former case to confirm the exchange by taking the absorption spectra of the two electrophoretic components of the recombined mixture.

Hb^+ A and HbCO C are not separable by electrophoresis, a result which is in accord with chemical data that the polypeptide chains of hemoglobins A and C differ by a total of four units of charge. Recombination of Hb^+ A and HbCO C results in two new components having mobilities lesser and greater, respectively, than the common mobility of the original species (FIGURE 3*a*). The net change can be explained by asymmetrical recombination

$$(\alpha_2^{A+}\beta_2^{A+})^{+4} + (\alpha_2^{A}\beta_2^{C})^{+4} = (\alpha_2^{A}\beta_2^{A+})^{+2} + (\alpha_2^{A+}\beta_2^{C})^{+6}.$$

The hybrid product of symmetrical recombination $(\alpha^A\alpha^{A+}\beta^{A+}\beta^C)^{+4}$ would not separate from Hb^+ A and HbCO C; however, symmetrical recombination would not account for the observed new components.

Each of the mixtures HbCO A-Hb^+S, HbCO A-Hb^+C, and HbCO S-Hb^+C resolves into two components before recombination and produces two additional components on recombination. The net reactions postulated to account for these results are

$$\alpha_2^{A}\beta_2^{A} + (\alpha_2^{A+}\beta_2^{S+})^{+6} = (\alpha_2^{A+}\beta_2^{A})^{+2} + (\alpha_2^{A}\beta_2^{S+})^{+4},$$

$$\alpha_2^{A}\beta_2^{A} + (\alpha_2^{A+}\beta_2^{C+})^{+8} = (\alpha_2^{A+}\beta_2^{A})^{+2} + (\alpha_2^{A}\beta_2^{C+})^{+6},$$

and

$$(\alpha_2^{A}\beta_2^{S})^{+2} + (\alpha_2^{A+}\beta_2^{C+})^{+8} = (\alpha_2^{A+}\beta_2^{S})^{+4} + (\alpha_2^{A}\beta_2^{C+})^{+6}.$$

FIGURE 3*b* shows electrophoretic patterns of untreated and recombined mixtures of HbCO S and Hb^+C. The appearance of only two new species in each of this group of experiments excludes exchange of single chains, and the absence of hybrid species $(\alpha^A\alpha^{A+}\beta^A\beta^{S+})^{+3}$, $(\alpha^A\alpha^{A+}\beta^A\beta^{C+})^{+4}$, and $(\alpha^A\alpha^{A+}\beta^S\beta^{C+})^{+5}$ again excludes symmetrical recombination.

Recombination with radioactively labelled hemoglobin. Net transfer of radioactivity from one electrophoretic or chromatographic component to the other occurs without formation of new components when a mixture of labeled hemoglobin of one type and unlabeled hemoglobin of another is recombined. After

glycine-2-C^{14}-labeled HbCO A was recombined with unlabeled HbCO S or C, HbCO A and the abnormal HbCO were isolated by starch-block electrophoresis. Counts of the specific activity of each fraction showed that about one-fourth of the activity of HbCO A had been transferred to the abnormal HbCO, a

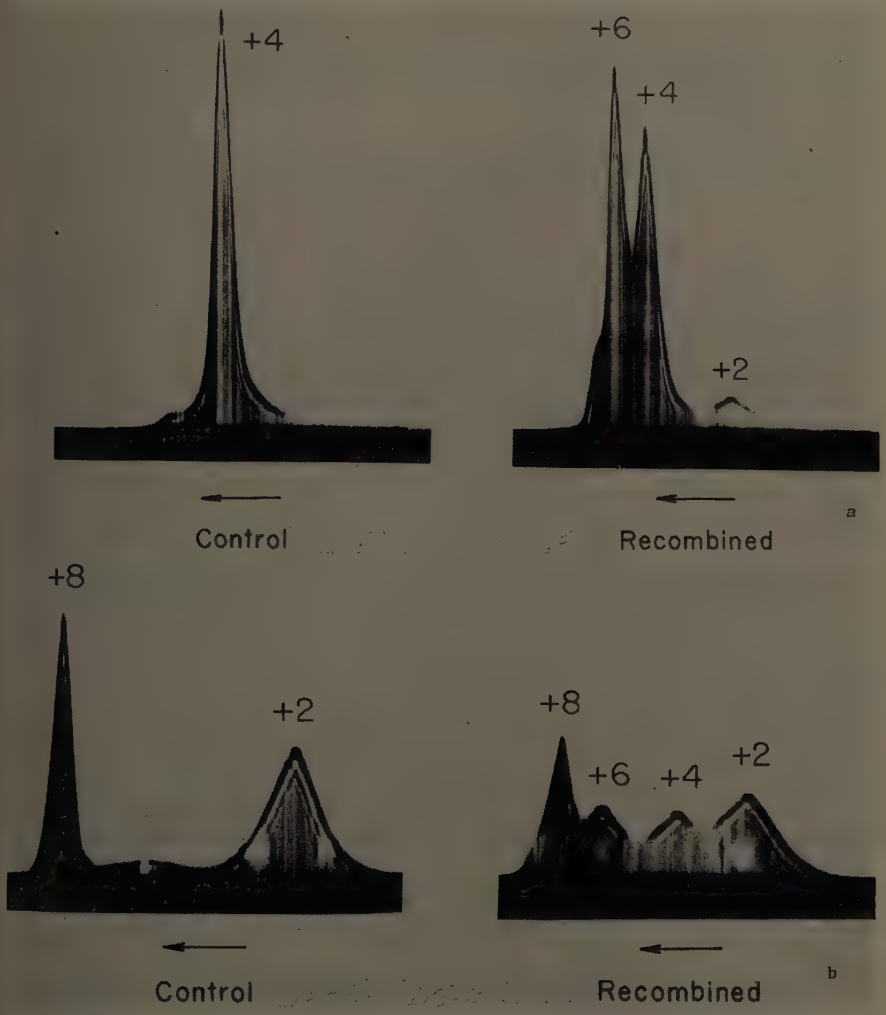


FIGURE 3*a*. Electrophoretic patterns of control and recombined mixtures of HbCO C and Hb^+ A.

FIGURE 3*b*. Electrophoretic patterns of control and recombined mixtures of HbCO S and Hb^+ C. From Singer and Itano.[3]

result consistent with asymmetrical recombination.[3] In the recombination of labeled HbCO A and unlabeled HbCO S the net reaction may be written

$$\alpha_2^{A*}\beta_2^{A*} + (\alpha_2^{A}\beta_2^{S})^{+2} = \alpha_2^{A}\beta_2^{A*} + (\alpha_2^{A*}\beta_2^{S})^{+2}.$$

Presumably HbCO A becomes a mixture of $\alpha_2^{A*}\beta_2^{A}$ and $\alpha_2^{A}\beta_2^{A*}$, and HbCO S

becomes a mixture of $(\alpha_2^{A*}\beta_2^{S})^{+2}$ and $(\alpha_2^{A}\beta_2^{S})^{+2}$. Recombination of unlabeled hemoglobins A and S, respectively, with hemoglobins S and A into which C^{14}-labeled leucine had been incorporated by reticulocytes resulted in an exchange of *N*-terminal peptides consistent with a β-chain defect in hemoglobin S.[4] Since separation of labeled and unlabeled molecules such as was observed with heme-labelled molecules does not occur, the same transfer of C^{14}-activity would be observed if the recombined mixture contained $\alpha^{A}\alpha^{A*}\beta_2^{A*}$ and $(\alpha^{A}\alpha^{A*}\beta_2^{S})^{+2}$. Therefore, exchange of single chains is not excluded by experiments in which radioactively labeled hemoglobins are used.

Applications

Detection of the aberrant chain of an electrophoretically abnormal hemoglobin. Asymmetrical dissociation to yield a common chain-pair was postulated when recombination of HbCO A, S, and C failed to produce electrophoretically distinct hybrid molecules.[2] New components do not appear when HbCO C and Hb^{+} S are recombined despite nonequivalence of α_2^{A} and α_2^{A+} because β_2^{C} and β_2^{S+} are equivalent in charge. On the other hand, new components appear when HbCO S and Hb^{+} C are recombined because of the difference in charge between α_2^{A} and α_2^{A+}, and between β_2^{S+} and β_2^{C+}. Thus, differences in charge increments of both chains are necessary to produce a net change in electrophoretic composition.

Inherited defects of human hemoglobin affect one or the other of its chains but not both;[7] consequently the altered chain of an abnormal hemoglobin can be identified by recombination with a previously characterized abnormal hemoglobin. The most readily available of the latter are hemoglobins S and C, which carry their chemical alterations in the β-chain. Recombination of HbCO I with HbCO S results in formation of HbCO A and a new component that migrates between A and I.[10] Net formation of the component with mobility of HbCO A without formation of any new component when HbCO I and HbCO C are recombined suggests that the species formed by combination of the β-chain of HbCO C and the α-chain of HbCO I has the same net charge as HbCO A. This interpretation is in accord with the observed abnormality in mobility of HbCO I, which is approximately equal and opposite to that of HbCO C. The two reactions involving HbCO I were deduced to be

$$(\alpha_2^{I}\beta_2^{A})^{-4} + (\alpha_2^{A}\beta_2^{S})^{+2} = \alpha_2^{A}\beta_2^{A} + (\alpha_2^{I}\alpha_2^{S})^{-2},$$

and

$$(\alpha_2^{I}\beta_2^{A})^{-4} + (\alpha_2^{A}\beta_2^{C})^{+4} = \alpha_2^{A}\beta_2^{A} + \alpha_2^{I}\beta_2^{C}.$$

Absence of observable change when HbCO A and HbCO I are recombined is consistent with presence of the same chain-pair β_2^{A} in both. Electrophoretic patterns of these experiments are shown in FIGURE 4, and compositions of control and recombined mixtures are given in TABLE 3. HbCO A was present in all the starting mixtures since homogeneous HbCO I was not available. HbCO A can exchange its common chain-pair β_2^{A} with HbCO I and α_2^{A} with HbCO S or C without net change in composition; therefore, the changes seen in FIGURE 4, (*d*) and (*f*), must arise from exchange of chain-pairs between HbCO I on the one hand and HbCO S and HbCO C, respectively, on the other.

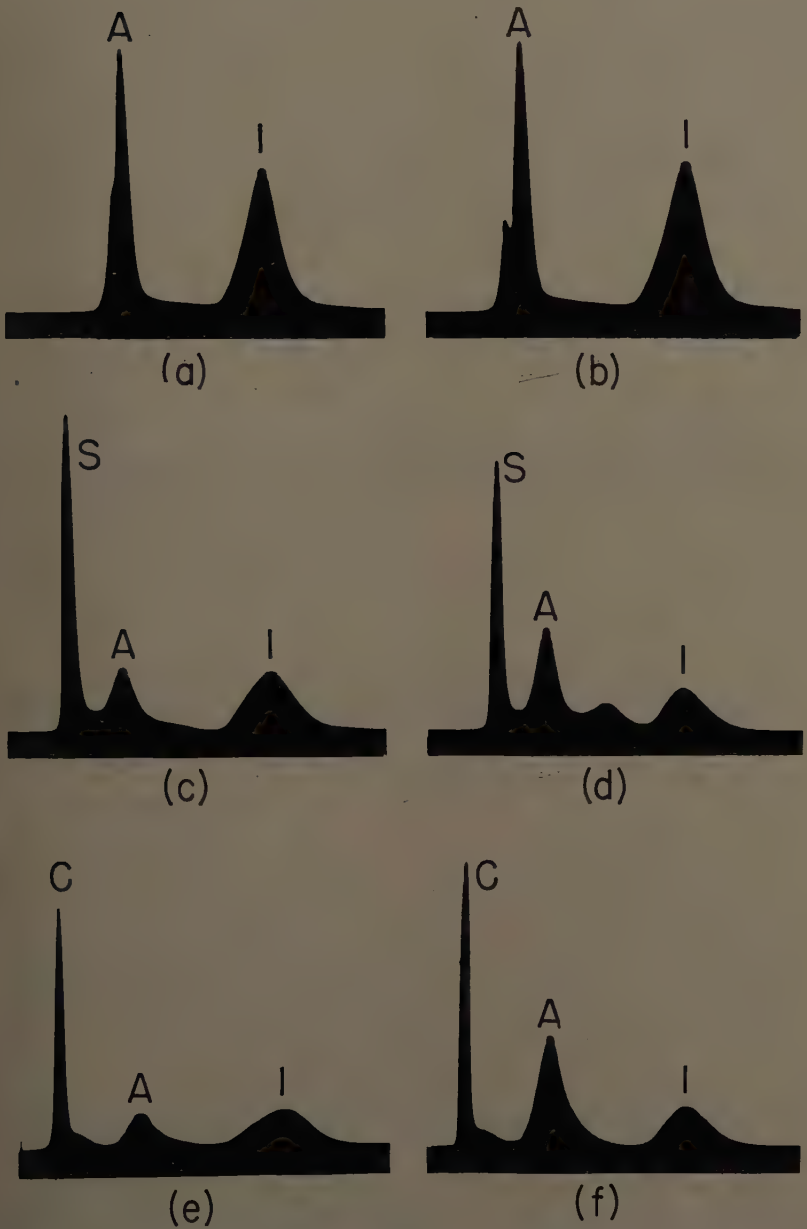


FIGURE 4. Electrophoretic patterns of control and recombined mixtures of HbCO I, S, and C. (*a*) A + I control; (*b*) A + I recombined; (*c*) A + I + S control; (*d*) A + I + S recombined; (*e*) A + I + C control; (*f*) A + I + C recombined. From Itano and Robinson.[10]

We have shown by asymmetrical recombination with HbCO S and C that hemoglobin Hopkins-2 is defective in the α-chain.[11] Individuals who have both hemoglobins S and Ho-2 have, in addition, a third electrophoretic component that has the mobility of hemoglobin A.[12] Since the abnormality in mobility of Ho-2 is approximately equal and opposite to that of S, we suggested that the third component is composed of two species with the same mobility, hemoglobin A and a species composed of the abnormal α-chain of Ho-2 and the abnormal β-chain of S.[11] This suggestion has been confirmed recently by recombination of a chromatographically isolated sample of the third component.[13] Two new components appeared in accordance with the reaction

$$\alpha_2^A\beta_2^A + \alpha_2^{Ho\text{-}2}\beta_2^S = (\alpha_2^A\beta_2^S)^{+2} + (\alpha_2^{Ho\text{-}2}\beta_2^A)^{-2}.$$

The appearance of two new electrophoretic components from one has a precedent in the recombination of HbCO-C and Hb$^+$A. Familial data reported

TABLE 3

COMPOSITION OF HEMOGLOBIN MIXTURES BEFORE AND AFTER RECOMBINATION[10]

Mixture		Electrophoretic composition: per cent of total				
		$(\alpha_2^I\beta_2^A)^{-4}$	$(\alpha_2^I\beta_2^S)^{-2}$	$\alpha_2^A\beta_2^A$	$(\alpha_2^A\beta_2^S)^{+2}$	$(\alpha_2^A\beta_2^C)^{+4}$
A + I	Control	53	—	47*	—	—
	Recombined	55	—	45*	—	—
A + I + S	Control	34	—	24*	42	—
	Recombined	23	11	32	34	—
A + I + C	Control†	36	—	20	—	44*
	Recombined	22	—	42‡	—	36*

* Includes 3 to 4 per cent of trailing shoulder.
† Re-run of original control; C peak too sharp in original for determination of area.
‡ Probably includes $\alpha_2^I\beta_2^C$.

by Smith and Torbert[12] that the genes that control hemoglobins S and Ho-2 are nonallelic suggest application of *in vitro* recombination as a genetic as well as a chemical test.

Recombination between hemoglobins of different species. Dissociation and reassociation of a mixture of human HbCO A and canine HbCO resulted in production of two additional components.[11] Since neither of the new components was observed when either hemoglobin was dissociated in the absence of the other, asymmetrical recombination was postulated:

$$\alpha_2^A\beta_2^A + \alpha_2^{can}\beta_2^{can} = \alpha_2^A\beta_2^{can} + \alpha_2^{can}\beta_2^A.$$

The same net reaction occurred in the recombination of alkali-dissociated mixtures of these hemoglobins (FIGURE 5).[14] TABLE 4 shows the composition of mixtures before and after exposure to *p*H 4.7 for 20 minutes to 22 hours.[11] The symbols α^{can} and β^{can} are used here to specify chain-pairs that are complementary to the β- and α-chains, respectively, of hemoglobin A since the *N*-terminal amino acid sequences of the chains of canine hemoglobin are not known. Structural analogues of the α- and β-chains in any hemoglobin that

recombines with human hemoglobin can be specified by this method. In a recent extension of this work we have recombined canine hemoglobin with abnormal human hemoglobins and deduced which of the new components carry the α^{can}- and β^{can}-chains.[15]

Test for dissociability of hemoglobin. No change was observed when bovine HbCO was acidified and neutralized in mixtures with human HbCO A and canine HbCO.[11] The possibility that these three species, like HbCO A, S, and C, contain an equivalent chain-pair is excluded by the appearance of new

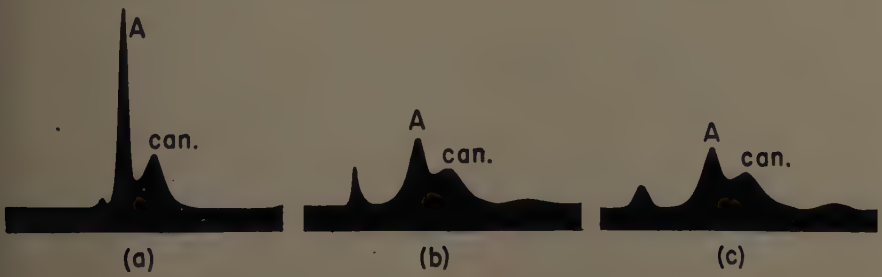


FIGURE 5. Recombination of human HbCO A with canine HbCO A. (*a*) Control mixture; (*b*) Recombined mixture after alkali-dissociation; (*c*) Recombined mixture after acid-dissociation. From Robinson and Itano.[14]

TABLE 4

COMPOSITION OF CANINE AND HUMAN HEMOGLOBIN MIXTURES BEFORE AND AFTER RECOMBINATION[11]

Dissociation time	Electrophoretic composition: per cent of total*			
	New slow	Canine	A	New fast
Control	—	41	59	—
20 minutes	10	36	44	10
40 minutes	11	39	39	11
4 hours	22	28	32	18
22 hours	31	24	21	24

* All samples contained 50 per cent each of human and canine hemoglobin before recombination. Percentages shown are probably anomalous because of low concentration of buffer ion.

components when human and canine HbCO are recombined. Other possible explanations are that the products of dissociation of bovine HbCO are not complementary to those of the other two or that bovine HbCO does not dissociate. We have shown that recombination of HbCO and Hb^+ from the same hemoglobin results in a single intermediate electrophoretic component (FIGURE 2*a*). However, such a component did not appear when a mixture of bovine HbCO and Hb^+ was acidified and neutralized. Assuming no structural limitation on association of subunits of the same hemoglobin to form an intermediate compound, we concluded that acidified bovine hemoglobin does not contain subunits, that is, that bovine hemoglobin does not dissociate at a *p*H and within a period of time sufficient to dissociate both human and canine hemo-

globins. The relative simplicity of the experiment and its interpretation suggests its use for testing other hemoglobins and other methods of dissociation.

Discussion

A three-dimensional model of horse hemoglobin has been obtained by Perutz by X-ray analysis.[16] The model shows how each polypeptide chain is folded and how these subunits are arranged in the molecule. The molecule has dyad symmetry, and the subunits are arranged tetrahedrally so that each subunit is in contact with the other three. The dyad is normal to the long axis of the molecule. The two types of chains are not folded identically; however, the chains resemble each other and sperm whale myoglobin in their general features. The hemes are partially buried on the surface of the molecule and are located approximately at the corners of a tetrahedron. It is unlikely that the major structural features of the hemoglobins we have studied differ substantially from those of horse hemoglobin. Complementarity of nonidentical chain pairs and the presence of one heme per chain, as deduced from recombination studies,[3] are in accord with the three-dimensional model. According to the model there are two possible ways in which dissociation into half-molecules $\alpha\beta$ can occur, and the products of the two modes of symmetrical dissociation are not complementary (FIGURE 1).

The most straightforward interpretation of the results of recombination after acid dissociation is that dissociation is asymmetrical. An alternative mechanism, that initial symmetrical dissociation is followed by exchange of single chains between half-molecules, is excluded by the experiments on mixture of HbCO and Hb^+, in which intermediate compounds with one or three ferrihemes did not appear. It is not clear why reversible dissociation in acid results in chain-pairs instead of single chains or why dissociation in acid is asymmetrical. Dissociation of horse hemoglobin into single chains occurs at higher acid concentration;[17] however, the process is accompanied by denaturation.

Species such as α_4^A, β_4^A, and $\beta_2^A\beta_2^S$ have not been found in recombined mixtures, although their presence in minute amounts is not excluded. Red cells doubly heterozygous for the genes that control synthesis of hemoglobins S and Ho-2 contain four different hemoglobins that result from association of four types of chains.[13] Occurrence of the four possible combinations of $\alpha_2\beta_2$ suggests that the chains are synthesized independently, and absence of tetrameric molecules, at least in amounts detectable by our methods, points to preferential association between heterologous chain-pairs *in vivo* as well as in *in vitro* recombination of dissociated mixtures. That this affinity cuts across species differences is shown by our *in vitro* results with human and canine hemoglobins. Hemoglobin H represents an exceptional case in which formation of a tetramer, β_4^A, is attributed to presence of an excess of β-chains in red cells.[18] Recombination of hemoglobin H with S results in formation of hemoglobin A and a tetramer, β_4^S, of the β-chain of hemoglobin S. Thus both β_2^A and β_2^S are complementary not only to α_2^A but also to homologous pairs of β-chains; on the other hand, the species $\beta_2^A\beta_2^S$ was not found in the recombined mixture.

Although it has been proposed that dissociation of hemoglobin in alkali is symmetrical,[19] net exchange of subunits on recombination of alkali-dissociated

mixtures is asymmetrical,[4,14] and hybrid molecules composed of a half-molecule $\alpha\beta$ of each of the original components do not form. The critical experiment of recombining HbCO and Hb^+ was unsuccessful because Hb^+ denatured too rapidly in alkali to permit reversible recombination. Structural features that might prevent association of two chains or half molecules that differ from each other by one amino acid residue have been discussed.[2] If no structural change is induced elsewhere than at the site of the amino acid difference, association can be prevented only if the difference is in the immediate vicinity of the dyad axis. Otherwise, it must be assumed that structural alterations are not localized to the site of the amino acid substitution. Neither possibility is supported by X-ray studies, which show that the chains are not in close apposition along the dyad and that hemoglobins A and S have identical X-ray diffraction patterns.[20]

Our limited experience with human, canine, and bovine hemoglobins is sufficient to point out both similarities and differences in structure that exist among hemoglobins. The subunits of human and canine hemoglobin are complementary to the extent that a completely recombined mixture contains approximately equal amounts of original components and recombinant species.[10] Bovine hemoglobin, which does not dissociate under conditions of acidity and time sufficient to dissociate human and canine hemoglobins, is also extraordinarily resistant to denaturation by alkali.[21] The latter finding may be contrasted with reports that bovine hemoglobin dissociates in urea, whereas human and canine hemoglobins do not.[22,23] Elucidation of structural characteristics that can account for the common properties of all hemoglobins and the unique properties of particular hemoglobins will very likely have to await knowledge of the spatial orientation of each constituent amino acid. Several groups of workers are now engaged in determining the amino acid sequence of hemoglobin, and complete structural determination of hemoglobin by correlation of their findings with the X-ray analyses of Perutz now appears attainable.

Summary

Human hemoglobin dissociates reversibly in acid into unlike chain-pairs α_2 and β_2. When a mixture of two molecular species is dissociated and then recombined by neutralization, exchange of chain-pairs occurs. The exchange can be detected with use of a mixture of unlabeled and labeled molecules or a mixture of molecules that are electrophoretically nonequivalent in both pairs of chains. Several applications of the process, which is termed asymmetrical recombination, have been suggested. The defective chain of an abnormal hemoglobin can be determined. Structural analogues of the human α- and β-chains can be specified in hemoglobins of other species. Whether a hemoglobin dissociates reversibly under a given set of conditions can be tested by subjecting a mixture of its carbonmonoxy- and ferri-forms to these conditions.

References

1. Itano, H. A., S. J. Singer & E. Robinson. 1959. *In* Biochemistry of Human Genetics. G. E. W. Wolstenholme & C. M. O'Connor, Eds. : 96. Churchill. London, England.
2. Itano, H. A. & S. J. Singer. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 522.
3. Singer, S. J. & H. A. Itano. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 174.

4. Vinograd, J. R., W. D. Hutchinson & W. A. Schroeder. 1959. J. Am. Chem. Soc. **81:** 3168.
5. Itano, H. A. & E. Robinson. 1958. Biochim. et Biophys. Acta. **29:** 545.
6. Rhinesmith, H. S., W. A. Schroeder & N. Martin. 1958. J. Am. Chem. Soc. **80:** 3358.
7. Hunt, J. A. & V. M. Ingram. 1959. *In* Biochemistry of Human Genetics. G. E. W. Wolstenholme & C. M. O'Connor, Eds. : 114. Churchill. London, England.
8. Field, E. O. & J. R. P. O'Brien. 1955. Biochem. J. **60:** 656.
9. Hasserodt, U. & J. Vinograd. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 12.
10. Itano, H. A. & E. Robinson. 1959. Nature. **183:** 1799.
11. Itano, H. A. & E. Robinson. 1959. Nature. **184:** 1468.
12. Smith, E. W. & J. V. Torbert. 1958. Bull. Johns Hopkins Hosp. **101:** 38.
13. Itano, H. A. & E. Robinson. Federation Proc. **19:** 193.
14. Robinson, E. & H. A. Itano. Nature. **185:** 547.
15. Robinson, E. & H. A. Itano. In preparation.
16. Perutz, M. F., M. G. Rossman, A. F. Cullis, H. Muirhead, G. Will & A. C. T. North. 1960. Nature. **185:** 416.
17. Reichmann, M. E. & J. R. Colvin. 1956. Can. J. Chem. **34:** 364.
18. Jones, R. T., W. A. Schroeder, J. E. Balog & J. R. Vinograd. 1959. J. Am. Chem. Soc. **81:** 3161.
19. Kendrew, J. C. 1959. Federation Proc. **18:** 740.
20. Perutz, M. F., A. M. Liquori & F. Eirich. 1951. Nature. **167:** 929.
21. Haurowitz, F., R. L. Hardin & M. Dicks. 1954. J. Phys. Chem. **58:** 103.
22. Wu, H. & E. F. Yang. 1932. Chinese J. Physiol. **6:** 51.
23. Gutter, F. J., H. A. Sober & E. A. Peterson. 1956. Arch. Biochem. and Biophys. **62:** 427.

Part III. Organic Synthesis of Peptides

STEPWISE SYNTHESIS OF PEPTIDES BY THE NITROPHENYL-ESTER METHOD

Miklos Bodanszky
The Squibb Institute for Medical Research, New Brunswick, N. J.

When, about sixty years ago, the classic studies of Emil Fischer[1] proved that proteins are built up from amino acids, the synthetic methods devised by him and by Theodor Curtius[2] were adequate enough to link amino acids to each other by peptide bonds. Thirty years later the introduction of the carbobenzoxy protecting group discovered by Bergman and Zervas[3] made it possible to synthesize simple peptides such as carnosin,[4] anserin,[5] and glutathion.[6] Ten years ago a number of new methods were developed that could all be classified as mixed anhydride methods (within this group can be included also the so-called amino group activating methods). The discovery of the mixed anhydride procedures[7a-e] was most opportune. Almost simultaneously, structures of relatively complicated peptides like oxytocin and the vasopressins were elucidated by du Vigneaud and his co-workers[8] and, with the new methods, the synthesis of these hormones[9a-b] could be achieved. More recently a number of interesting and important natural products of peptide structure have been synthesized, such as the antibiotic Gamicidin S,[10] the substances raising blood pressure, angiotensins[11a-b], still later, α- and β-melanotropins.[12a-c] It was essentially the aforementioned synthetic methods that were applied in these syntheses. However the synchronization of progress in analysis and synthesis is no longer present in protein chemistry. Recent developments in the study of amino acid composition and sequence in proteins has led to the fascinating elucidation of structures as complex as insulin,[13] ACTH,[14] glucagon,[15] and even of ribonuclease,[16] thereby far outstripping our synthetic capacity. The present state of synthesis seems to be unable to compete with the methods of analysis. Consequently not only the practical preparation of these hormones and enzymes is impossible at the moment, but even the proof of their structure by synthesis still lies in the future.

Above a certain molecular size in a peptide, the difficulties of synthesis arise from several factors. There is the problem of the conformation of peptides, that is, the secondary structures due to such factors as hydrogen bonding, folding, and helix structure that may require special planning not known as yet. Another problem arises where, in a peptide to be synthesized, there is more than one disulfide bridge. The formation of a ring with an S—S bridge by oxidation of two SH groups to a disulfide—so brilliantly achieved by du Vigneaud and his co-workers[17]—is not practical if more than two SH groups are to be linked.

In any event, apart from these special problems, the more general problem of peptide bond formation is only partially solved by the presently available methods that have been successfully applied in the synthesis of smaller molecules. Building a long chain involves many synthetic steps and, unless all of these steps can be performed with excellent yields, unusually large amounts of starting materials have to be used to obtain a decent quantity at the comple-

tion of the synthesis. In this respect, it is also very desirable to employ only methods that lead to a single product, thereby avoiding the need of purifying intermediates. Since purification usually involves such procedures as chromatography, electrophoresis, or countercurrent distribution, and, useful though such methods may be in the isolation of natural products, they are stumbling blocks if applied several times in a synthetic procedure. They hinder the synthesis in two ways: (1) by limiting the amounts that can be handled, and (2) by consuming time in their operation.

From the above considerations, it is evident that the mixed-anhydride methods are not always desirable in the synthesis of long peptide chains. Invariably with these methods there are formed two products as shown in FIGURE 1, which illustrates acylation of an amine with the mixed anhydride of an acylamino acid and isovaleric acid.

Although the amide (1) shown above the line is the main product, it is not the only one, and small amounts of the acylamino acid (3) shown below the dotted line can always be isolated from the reaction mixture, indicating the formation of a second neutral product (4). Of course, while the second, un-

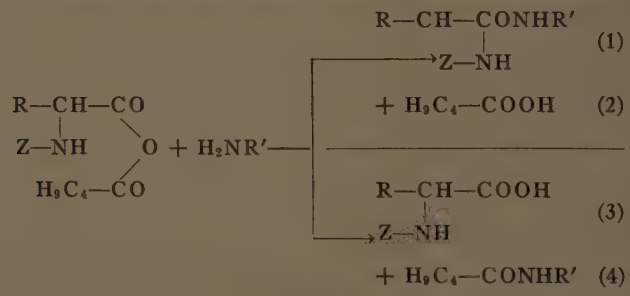


FIGURE 1.

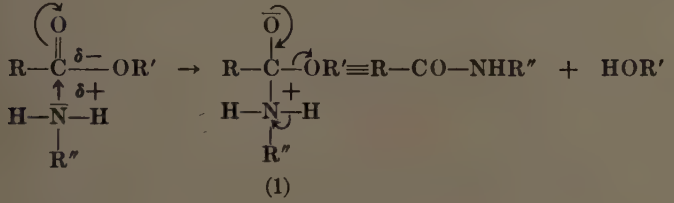
desirable acylation with well-chosen "partner" acids occurs only to a minor extent, it is insignificant only in the synthesis of a dipeptide or a tripeptide, where crystallization of the main product leaves the undesired amide in the mother liquor. When the amino component H_2NR' in the above reaction is a peptide of considerable length and is lengthened by one amino acid, the two resulting amides do not differ much from each other; and their separation by crystallization may not be easy, especially since the crystallization of the main product is impeded by the presence of a contaminant of almost identical structure.

This same shortcoming is possessed, of course, also by those amino group activating methods that give rise to mixed anhydrides during reaction. Side reactions leading to undesirable by-products are associated with other methods of peptide synthesis too, such as the azide, the acid chloride, and even the carbodiimide procedure.[18]

For these reasons, the need for a more unequivocal procedure became evident. Among the possibilities considered, the aminolysis of esters recommended itself for compelling reasons: (1) The only by-product of the reaction is the liberated alcohol component; (2) acylation occurs only on NH_2 groups but not on SH or

OH groups as in the mixed anhydride methods. In 1954, therefore, we started our study in synthesis of peptides with esters of acylamino acids.

As is well-known, Curtius had used the self-condensation of methyl esters of amino acids and peptides for the preparation of longer-chain peptide derivatives.[19] It is obvious, however, that such an uncontrolled self-condensation leads to a number of products. This could be avoided by the use of protective groups but, since methyl esters are comparatively inert, rather drastic heating—without a diluent—is required to effect the condensation. In our search for more reactive esters we found that Gordon *et al.*[20a-b] had investigated already the general rules for ester aminolysis. These workers found that the rate of this reaction for unsubstituted, saturated, aliphatic alcohol components is greatest with methyl esters and decreases as the length of the chains is increased. Thus, compared to the methyl ester, the rate for the ethyl ester is about one half and still less with propyl and other esters. These authors also mention that in aminolysis vinyl esters are several hundred times, and phenyl esters about a thousand times more active than methyl esters. Apparently, the



(1)

FIGURE 2.

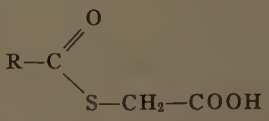
electrophilic character of the carbonyl carbon atom in such esters is enhanced by the electron-attracting alcohol component.

The formation, therefore, of an intermediate product [FIGURE 2 (1)] by the nucleophilic attack of the amine is facilitated by the positive character of the carbonyl carbon atom in these active esters. In this regard, it became obvious that the thiophenyl-ester procedure of Wieland[21] is essentially an active-ester method, even though this investigator called the thiophenyl esters of acylamino acids mixed anhydrides, since thiophenol can be regarded an acid. Wieland's considerations had started from analogues in biochemistry, the thiol esters. It can be seen now, however, that it is the fact that they are phenyl esters rather than thiol esters that is the main reason for their reactivity, while the sulfur atom with its larger shell of electrons is more efficient in transferring the inductive effect towards the carbonyl carbon atom than oxygen. Nonetheless, by rendering the alcohol component still more electron-attracting by substitution with a nitro group the resulting nitrophenyl ester exhibits far better reactivity than the corresponding thiophenyl ester. For instance no refluxing of the reaction mixture is necessary as in the case of thiophenyl esters, and the acylation reaction even proceeds smoothly at room temperature.

Following soon after our first publication[22a-d] on this method of peptide synthesis, Kenner, Farrington, and Turner[23] independently reported their studies leading to the development of acylamino acid ester of *p*-nitrothiophenol.

While the reaction of these esters is even faster, this advantage is offset by the troublesome problem of the removal of the liberated nitrothiophenol.

It should be mentioned here that another group led by R. Schwyzer has also adapted the use of active esters.[24a-d] In continuation of Wieland's studies on thiolesters this Swiss group proposed acylamino acid esters using thioglycolic acid as the alcohol component.



Soon they dropped the sulfur atom and used other methyl esters (O-methyl esters) with substituents of more pronounced electron attraction. From a series of thorough investigations they concluded that, from a practical point of view, the esters of cyanmethyl alcohol (or rather glycolic acid nitrile) surpassed any other. A few years later, however, in 1957, in the synthesis of Gramicidin S,[10] Schwyzer and his co-workers adopted our nitrophenyl ester method in place of their own procedures. In the preparation of the active esters[25] they

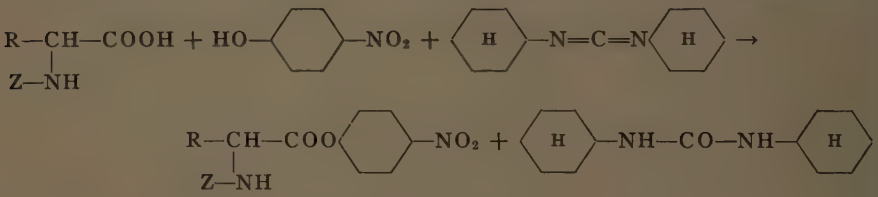


FIGURE 3.

applied an exchange reaction between the acylamino acid and *p*-nitrophenyl phosphite or sulfite, a method very similar to that used earlier by Kenner, *et al.*,[23] for the same purpose. This exchange reaction proceeds via mixed anhydrides[25] and, therefore, it does not differ basically from the way Wieland prepared his thiophenyl esters or from our own preparation of the nitrophenyl esters. The appealing features of the nitrophenyl ester method explain why, from two more laboratories working independently, the same new procedure was suggested for the esterification reaction. Thus, as an additional expedient in this method, Rothe[26] in Germany and Russel and Elliot[27] in England have used dicyclohexylcarbodiimide as a condensing agent between the protected amino acid and *p*-nitrophenol (see FIGURE 3).

Undoubtedly one advantage in this procedure is the fact that the condensing agent, the diimide, is commercially available; another is easy removal of the urea derivative. Having adopted this modification, we generally use ethyl acetate as a solvent; and, following the removal of the urea derivative, the evaporation of the solvent, as a rule, leaves a crystalline residue from which the analytically pure ester is obtained after one recrystallization from ethanol.*[28]

* Under these conditions the method is definitely superior to the exchange reaction with nitrophenyl sulfites or phosphites. For example, *p*-nitrophenyl carbobenzoxy glycinate has been reported from Schwyzer's laboratory several times having a m.p. 124 to 125° C. whereas with the aid of the carbodiimide invariably a product melting at 128 to 128.5° C. is obtained.

At this point it is appropriate to mention that the nitrophenyl esters of acylamino acids are remarkably stable compounds with unlimited shelf life. Moreover, they exhibit excellent crystallization properties. Even the esters of carbobenzoxy-L-leucine and of carbobenzoxy-L-isoleucine crystallize readily even though their precursor acids have been obtained, so far, only as oils. The nitrophenyl esters, as might be expected, are sensitive towards alkaline hydrolysis yet very stable against acids. A great advantage in their use is that not only their chemical purity but also their optical purity can be ascertained easily. Under the conditions of peptide synthesis they are not racemized. In some recent experiments by J. E. Stouffer, D. Jarvis, and V. du Vigneaud (unpublished observations), however, it was found that on long standing in dimethyl formamide the values of optical rotation slowly diminish, the rate of racemization varying with derivatives of different amino acids. The process, probably caused by basic impurities in the solvent, is markedly fast in the case of the ester of N-carbobenzoxy-S-benzyl-L-cysteine; it is also catalyzed by the presence of a small amount of a tertiary base as triethylamine. Even this most sensitive ester is fully stable in acetic acid and also in dimethylformamide containing one per cent acetic acid. As a matter of fact, we believe that probably also other active derivatives of carbobenzoxy amino acids are subject to racemization if kept under similar conditions, that is, in the presence of a tertiary base and without an amine that could be acylated. The stability of the carbobenzoxy amino acid nitrophenyl esters makes it possible to observe this slow process. With the aid of active esters the rate of racemization of different amino acids and the protective power of different acyl groups can be studied. It can be concluded then that the protecting capacity of carbobenzoxy groups against racemization is not an absolute but only a relative one. Nonetheless, this is the most valuable and reliable protective measure in the synthetic work beset with the constant danger of racemization, and peptides with excellent optical purity were obtained even with the above mentioned ester of carbobenzoxy-benzyl-L-cysteine, since the aminolysis reaction is several magnitudes faster than the rate of racemization.

At this point we should return to our original problem, the building of long peptide chains. This can be done in three different ways: (1) Stepwise synthesis, from the C-terminal end of the chain, a method involving a maximum number of steps and continuously diminishing yield when calculated on the C-terminal amino acid; (2) Stepwise synthesis starting from the N-terminal amino acid, a method that involves a minimum number of steps, but also with decreasing yield if calculated for the *N*-terminal acid; (3) Linking together of smaller peptide chains by one of the aforementioned procedures. Obviously this last method offers the advantage of the minimum number of steps and a proportioned loss for all the amino acids involved.

If the question of racemization could be eliminated, the last mentioned method or, eventually, the second should be used, but definitely not the first. Racemization, however, is a factor that can not be ignored. It seems to intrude again and again, lessening the purity of the products in peptide synthesis. Against this, two efficient defensive actions can be taken: (1) the application of the azide method, and (2) the use of carbobenzoxy amino acids (not carbo-

benzoxy peptides) as acid components.*[29] Usually the combination of these two precautions is applied: smaller peptide intermediates are built up stepwise from protected amino acids, and the peptides thus obtained are linked to each other by the azide method. Unfortunately, the last mentioned process, even if free from racemization, leads to a number of by-products other than diastereoisomers, necessitating a careful purification of the coupling product. Moreover, technical difficulties emerge in the azide synthesis as the number of amino acids increases. For example, the hydrazides obtained from longer chains are quite insoluble, the reaction rate of the azides is very slow, and so forth.

The stepwise synthesis starting from the *N*-terminal acid is eliminated at the outset because of its inability to form optically pure products. The method of

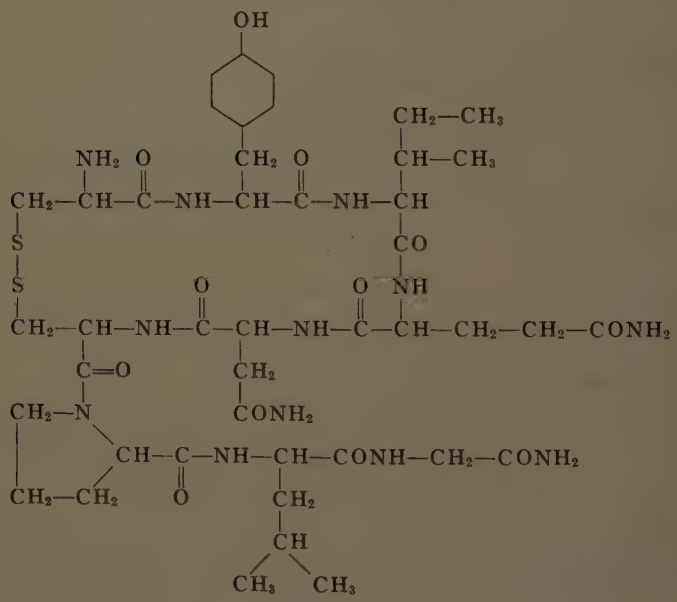


FIGURE 4. Oxytocin.

linking peptides together by the azide procedure also has its drawbacks. Can one find recourse in the method that seemed to be the least attractive in our original considerations? We felt that this question could be answered in the affirmative if the synthetic methods used gave excellent yields and led to a single product. If these conditions were fulfilled, the number of steps would play a less important role and, since only protected amino acids and not protected peptides were used, the protective power of the carbobenzoxy group against racemization would be utilized. With du Vigneaud we decided to test the practicability of this manner of synthesis in a resynthesis of oxytocin. Oxytocin (FIGURE 4) was chosen as the best model not only because it offered opportunity to compare the results of the new synthesis with several earlier ones, but

* Of course other protective groups can also protect against racemization, but not all acyl groups are satisfactory in this sense; for example, acetyl does not protect at all.

also because a direct comparison with the pure natural hormone for biological activity in different tests gives direct evidence of the grade of purity of the synthetic product.

Following the original epoch making synthesis of oxytocin by du Vigneaud and his co-workers,[9] the subsequent preparations of the hormone[30-35] have involved intermediate-size peptides linked together by different methods to give nonapeptide derivatives differing only in respect to the protecting groups. After the removal of these protective groups, oxidation led to oxytocin (FIGURE 5).

Since only the three syntheses at Cornell were carried out to the isolation of the pure hormone, it is not possible to compare the yields in all the different procedures. A rough calculation shows, however, that the earlier syntheses resulted in over-all yields of under 5 per cent;[34] one was performed with about 10 per cent, and the stepwise synthesis of oxytocin[36] gave a 30 per cent over-all yield. In addition, when judged from the standpoints of melting points or of yield on biological activity, the purity of the protected nonapeptide was also much superior to those reported in the earlier preparations.

Another comparison of different methods leading to the same compound is the synthesis of 2-phenylalanine oxytocin,[37] a biologically active analogue of the hormone where phenylalanine replaces tyrosine. This hormone analogue was prepared simultaneously and independently also by Jaquenoud and Boissonnas.[38] In this case, the stepwise synthesis by the nitrophenyl-ester method proved to be superior both with respect to yields and to the purity of the protected nonapeptide.

In more recent studies,[39] the S,S'-dibenzyl derivative of the reduced form of oxytocin was converted into a decapeptide derivative by the action of the *p*-nitrophenyl ester of carbobenzoxyglycine. The lengthening of the nonapeptide chain by one amino acid gave a quantitative yield of the protected decapeptide from which, in turn, N-glycyloxytocin was obtained. Its interesting pharmacological properties will be discussed elsewhere.

Another successful application of the stepwise synthesis with nitrophenyl esters is represented in the preparation of lysine vasopressin (FIGURE 6). In this study,[40] all the protected intermediates were obtained in crystalline form without any preliminary purification, and the over-all yield in the nine peptide bond-forming steps leading to a protected nonapeptide is about 50 per cent. In the conversion of this intermediate into the biologically active hormone, the presence of only insignificant amounts of inactive by-products could be demonstrated. Thus, it can be seen that the stepwise synthesis by the nitrophenyl-ester method seems to be capable of producing practically pure peptides in satisfactory yields even after a considerable number of synthetic steps.

Let us consider finally one more rather technical aspect of this procedure. Here, the lengthening of the already existing peptide chain by one amino acid was done in a rather mechanical manner with a few standard operations. Thus, to the solution of the amino component (the peptide chain with a free amino group), the nitrophenyl ester of the carbobenzoxy-amino acid to be attached is added in some excess. This excess is applied to make certain that the valuable amino component is completely utilized. For the same reason, although the product usually crystallizes from the solution within a few hours,

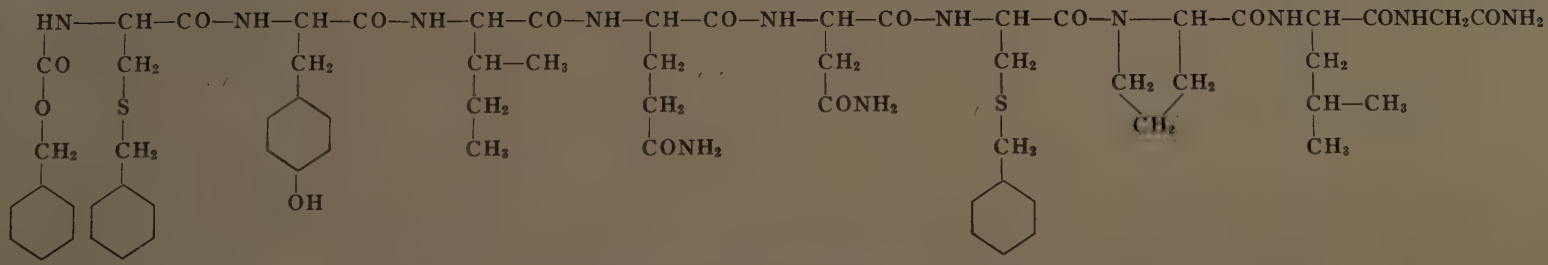


FIGURE 5. A protected nonapeptide intermediate in the synthesis of oxytocin.

the reaction mixture is left to stand at room temperature for one or two days. Addition of ethyl acetate to the reaction mixture completes the separation of the desired product, but leaves in solution both the excess of the starting nitrophenyl ester and the nitrophenol liberated during the reaction. Washing with liberal quantities of ethyl acetate renders the product analytically pure. Even if this picture is somewhat over-simplified, it projects another picture, that of the peptide synthesis of the future. With commercially available active esters of protected amino acids and with simple operations such as depicted above, peptide synthesis can be performed by mechanical devices. This view may be fantastic at the moment, but it is not more fantastic than the automatic

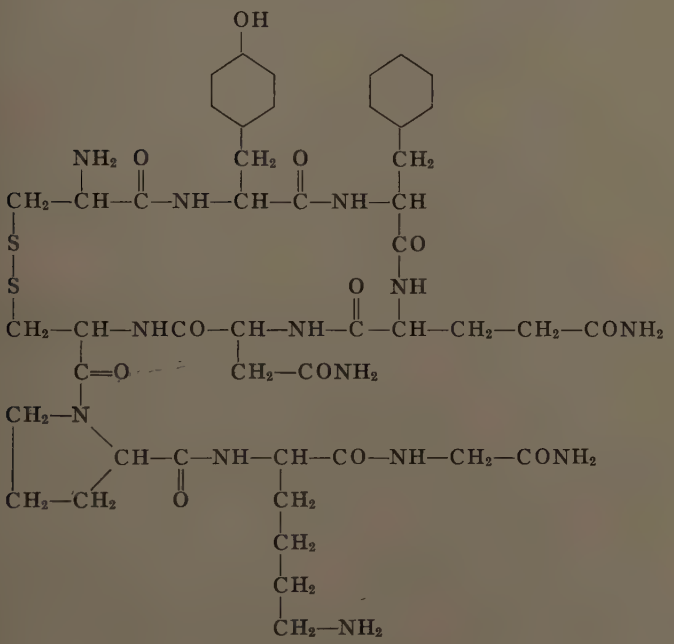


FIGURE 6. Lysine vasopressin.

analysis of proteins and peptides for amino acid content would have seemed, let us say, 20 years ago. It certainly seems to be probable that, for the synthesis of really long chains, simplification and mechanization of the procedure is as essential as in the analysis of proteins. The development of simple and mechanically applicable methods can, sometime from now, close the recently opened gap between analysis and synthesis in this field of natural products.

References

1. FISCHER, E. 1906. Untersuchungen über Aminosäure, Polypeptide und Proteine. Springer. Berlin, Germany.
2. CURTIUS, T. 1904. J. prakt. Chem. **70:** 57. 1902. Ber. D. Chem. G. **35:** 3226.
3. BERGMANN, M. & L. ZERVAS. 1932. *Ibid.* **65:** 1192.
4. SIFFERD, R. H. & V. DU VIGNEAUD. 1935. J. Biol. Chem. **108:** 753.
5. BEHRENS, O. K. & V. DU VIGNEAUD. 1937. J. Biol. Chem. **120:** 517.

6. DU VIGNEAUD, V. & G. L. MILLER. 1936. J. Biol. Chem. **116:** 469.
7*a*. WIELAND, T., W. KERN & R. SEHRING. 1950. Ann. Chem. Liebig. **569:** 117.
7*b*. WIELAND, T. & R. SEHRING. 1950. *Ibid*. **569:** 122.
7*c*. VAUGHAN, J. R., JR. 1951. J. Am. Chem. Soc. **73:** 3547.
7*d*. VAUGHAN, J. R., JR. & R. L. OSATO. 1951. J. Am. Chem. Soc. **73:** 5553.
7*e*. BOISSONNAS, R. A. 1951. Helv. Chim. Acta. **34:** 874.
8. DU VIGNEAUD, V. 1955. XIVth International Congress of Pure and Applied Chemistry. Experientia Supplementum. **2:** 9. Zurich, Switzerland.
9*a*. DU VIGNEAUD, V., C. RESSLER, J. M. SWAN, C. W. ROBERTS, P. G. KATSOYANNIS & S. GORDON. 1953. J. Am Chem. Soc. **75:** 4879.
9*b*. DU VIGNEAUD, V., D. T. GISH & P. G. KATSOYANNIS. 1954. J. Am. Chem. Soc. **76:** 4751.
10. SCHWYZER, R. & P. SIEBER. 1956. Angew. Chem. **68:** 518.
11*a*. BUMPUS, F. M., H. SCHWARZ & I. H. PAGE. 1957. Science. **125:** 886.
11*b*. SCHWYZER, R., B. ISELIN, H. KAPPELER, B. RINIKER, W. RITTEL & H. ZUBER. 1957. Helv. Chim. Acta. **41:** 1273.
12*a*. GUTTMANN, S. & R. A. BOISSONNAS. 1959. Helv. Chim. Acta. **42:** 1257.
12*b*. SCHWYZER, R., H. KAPPELER, B. ISELIN, W. RITTEL & H. ZUBER. 1959. Helv. Chim. Acta. **42:** 1703.
12*c*. HOFMANN, K., M. E. WOOLNER, H. YAJIMA, G. SPÜHLER, T. A. THOMPSON & E. T. SCHWARZ. 1958. J. Am. Chem. Soc. **80:** 6458.
13. SANGER, F. 1955. Proc. Intern. Wool Textile Research Conf. **C:** 396. Melbourne, Australia.
14. LI, C. H. 1956. Advances in Protein Chem. **11:** 101.
15. BEHRENS, O. K. & W. W. BROMER. 1958. Vitamins and Hormones. **16:** 263.
16. HIRS, C. H. W., W. H. STEIN & S. MOORE. 1958. Symposium on Protein Structure. **:** 211. Methuen. London, England.
17. LORING, H. S. & V. DU VIGNEAUD. 1935. J. Biol. Chem. **111:** 385.
18. SHEEHAN, J. C. & G. P. HESS. 1955. J. Am. Chem. Soc. **77:** 1067.
19. CURTIUS, T. 1904. J. prakt. Chem. **70:** 54. Ber. D. Chem. G. **37:** 1284.
20*a*. GORDON, M., J. G. MILLER & A. R. DAY. 1949. J. Am. Chem. Soc. **70:** 1946.
20*b*. GORDON, M., J. G. MILLER & A. R. DAY. 1949. J. Am. Chem. Soc. **71:** 1245.
21. WIELAND, T., W. SCHAFER & E. BOKELMANN. 1951. Ann. Chem. Liebig. **573:** 99.
22*a*. BODANSZKY, M. 1955. Nature. **175:** 685.
22*b*. BODANSZKY, M., M. SZELKE, E. TOMORKENY & E. WEISZ. 1955. Chem. & Ind. **:** 1517.
22*c*. BODANSZKY, M. 1957. Acta. Chim. Hung. **10:** 335.
22*d*. BODANSZKY, M., M. SZELKE, E. TOMORKENY & E. WEISZ. 1957. Acta. Chim. Acad. Sci. Hung. **11:** 179.
23. FARRINGTON, J. A., G. W. KENNER & J. M. TURNER. 1955. Chem. & Ind. **:** 601.
24*a*. SCHWYZER, R. 1954. Helv. Chim. Acta. **37:** 674.
24*b*. SCHWYZER, R., B. ISELIN & M. FEURER. 1955. Helv. Chim. Acta. **38:** 69.
24*c*. SCHWYZER, R., M. FEURER, B. ISELIN & H. KAGI. 1955. Helv. Chim. Acta. **38:** 80.
24*d*. SCHWYZER, R., M. FEURER & B. ISELIN. 1955. Helv. Chim. Acta. **38:** 83.
25. ISELIN, B., W. RITTEL, P. SIEBER & R. SCHWYZER. 1957. Helv. Chim. Acta. **40:** 373.
26. ROTHE, M. & F. W. KUNITZ. 1957. Ann. Chem. Liebig. **609:** 88.
27. ELLIOT, D. F. & D. W. RUSSEL. 1957. Biochem. J. **66:** 49P.
28. BODANSZKY, M. Biochem. Prep. In preparation.
29. NORTH, M. B. & G. T. YOUNG. 1955. Chem. & Ind. **:** 1597.
30. BOISSONNAS, R. A., S. GUTTMANN, P. A. JAQUENOUD & J. P. WALLER. 1955. Helv. Chim. Acta. **38:** 1491.
31. BODANSZKY, M., L. GYERMEK & G. FEKETE. 1955. Experientia. **11:** 238.
32*a*. RUDINGER, J., J. HONZL & M. ZAORAL. 1956. Chem. listy. **50:** 288.
32*b*. RUDINGER, J., J. HONZL & M. ZAORAL. 1956. Coll. Czechoslov. Chem. Commun. **21:** 202.
33. VELLUZ, L., G. AMIARD, V. BARTOS, B. GOFFINET & R. HEYMES. 1956. Bull. soc. chim. France. **:** 1464.
34. BODANSZKY, M. & V. DU VIGNEAUD. 1959. J. Am. Chem. Soc. **81:** 2504.
35. BEYERMAN, H. C., J. S. BOUTEKOE & A. C. KOCH. 1959. Rec. Trav. Chim. **78:** 935.
36. BODANSZKY, M. & V. DU VIGNEAUD. 1959. Nature. **183:** 1324. J. Am. Chem. Soc. **81:** 5688.
37. BODANSZKY, M. & V. DU VIGNEAUD. 1959. J. A. Chem. Soc. **81:** 1258, 6072.
38. JAQUENOUD, P. A. & R. A. BOISSONNAS. 1959. Helv. Chim. Acta. **42:** 788.
39. FITT, P., M. BODANSZKY & V. DU VIGNEAUD. Proc. Soc. Exptl. Biol. Med. In preparation.
40. BODANSZKY, M., J. MEIENHOFER & V. DU VIGNEAUD. J. Am. Chem. Soc. **82.** In press.

ACTIVATED CYCLIC DERIVATIVES OF AMINO ACIDS IN PEPTIDE SYNTHESIS

John C. Sheehan

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Mass.

Many amino acids of interest to the chemist and biochemist contain, in addition to a carboxyl and an amine function, "extra" functional groups. These include both aliphatic and aromatic type hydroxyl groups, a thiol function, or a potentially reactive ring system. FIGURE 1 illustrates five important amino acids that have reactive functions that may interfere in synthetic processes.

The organic chemist interested in synthetic processes has usually dealt with the problems presented by extra functional groups in amino acids in either of

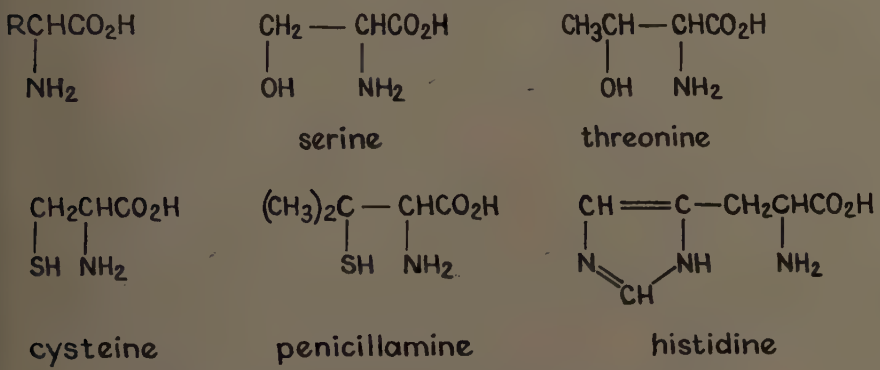


FIGURE 1. Typical amino acids containing extra functional groups.

two ways. First, the extra function is blocked as a separate operation, and then the protective group is removed after the desired peptide linkage has been formed. Examples of this type are the successful use of the *S*-benzyl thio-ether by Wood and du Vigneaud[1] in the formation of cysteinyl peptides and more recently the use of an *N*-formyl thiazolidine intermediate, also for the synthesis of cysteinyl derivatives.[2] A second scheme is to use an amide-forming reagent so selective that the extra function is not affected. An example is the synthesis of serine and threonine peptides by use of the dicyclohexylcarbodiimide reagent.[3]

However, it seemed to my colleagues and me that there might be still another possibility for dealing with the troublesome extra function. It might be possible to block the extra function and to activate the carboxyl group at the same time; in that way, a liability would be converted into an asset. One example of this approach is illustrated in FIGURE 2. *N*-trityl serine is converted by diisopropylcarbodiimide into a crystalline β-lactone. In common with other β-lactones, the carboxyl function is very reactive and on treatment with an amino acid ester the peptide derivative is formed as illustrated in FIGURE 3.

The condensation of *N*-trityl serine β-lactone with L-alanine methyl ester gives a 66 per cent yield of the peptide ester with no evidence of racemization.

Another example of the use of an activated cyclic derivative for peptide synthesis is outlined in FIGURE 4. *N*-(*p*-nitrocarbobenzyloxy)-L-histidine is cyclized in good yield to a crystalline activated derivative, which condenses in good yield with amino acid esters to form the corresponding peptide derivatives.

A third example of an activated cyclic derivative is provided by the class of

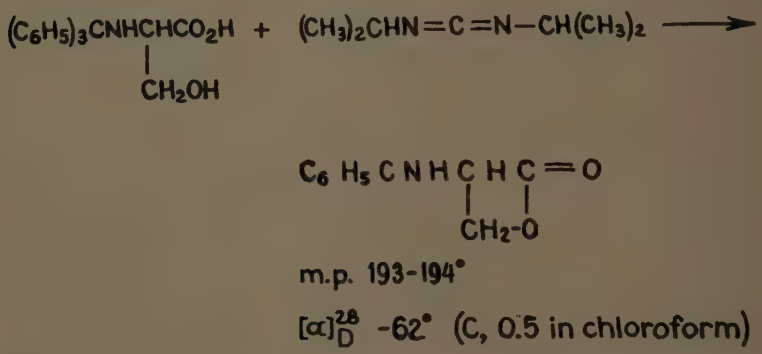


FIGURE 2. A serine β-lactone. Reproduced by permission of the *Journal of the American Chemical Society*.

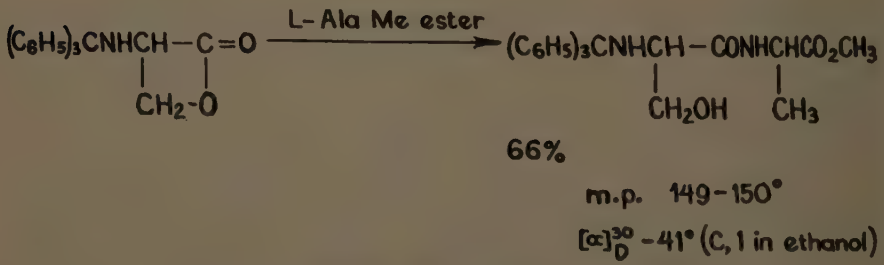


FIGURE 3. Peptide synthesis through a serine β-lactone.

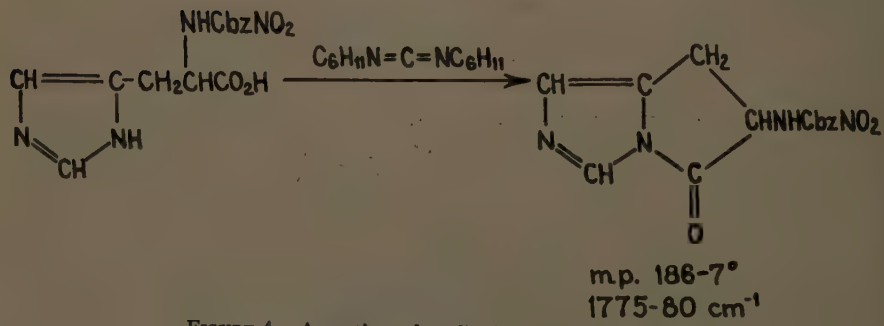


FIGURE 4. An activated cyclic derivative of histidine.

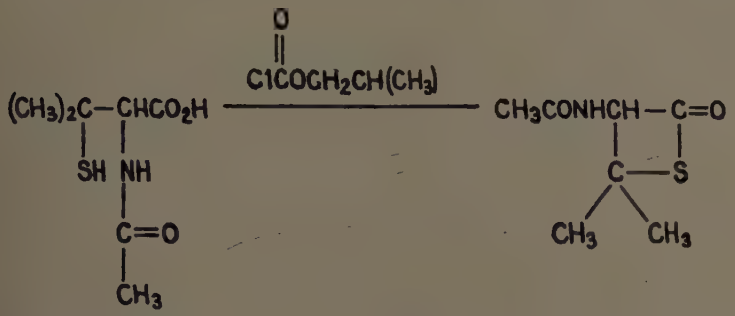


FIGURE 5. A β-thiolactone from *N*-acetylpenicillamine. Reproduced by permission of *Izvestiya Akademii Nauk S.S.S.R.*

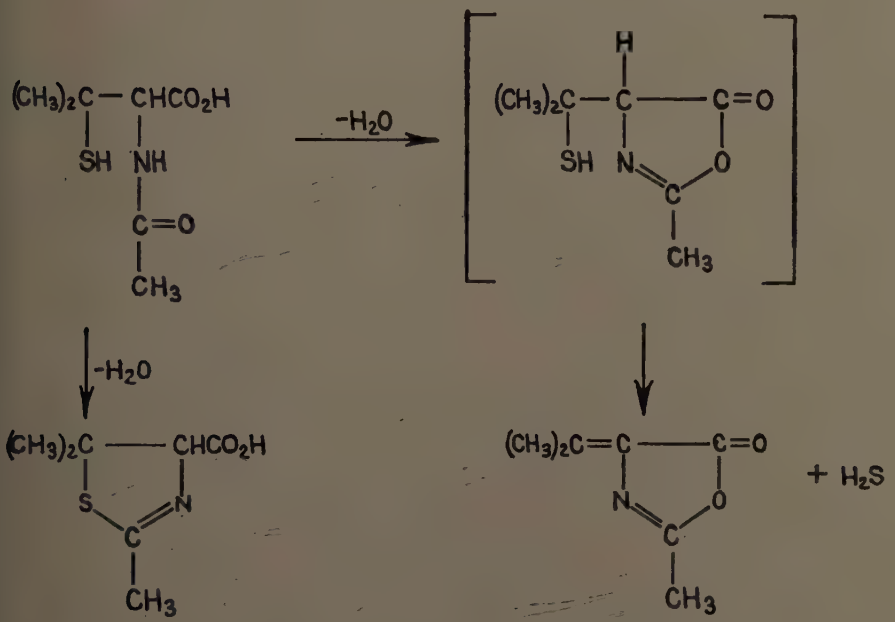


FIGURE 6. Possible modes for the dehydration of *N*-acetylpenicillamine.

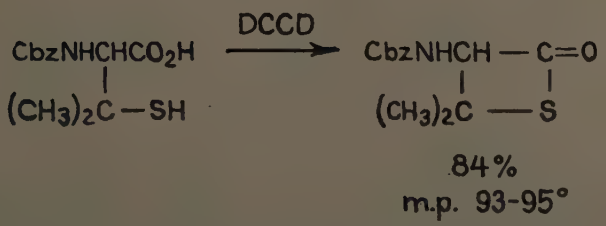


FIGURE 7. β-thiolactone derived from β,β-dimethylcysteine.

β-thiolactones. In 1955 a group of Soviet workers[4] reported the cyclization of *N*-acetyl-β,β-dimethylcysteine to the corresponding β-thiolactone, as illustrated in FIGURE 5. The structural evidence given by the Soviet group does not conclusively eliminate alternative structures, including a thiazoline or an azalactone, as illustrated in FIGURE 6. Knunyants also prepared *N*-acyl di- and tripeptide derivatives of DL-β,β-dimethylcysteine. However, the *N*-blocking acyl groups that he used (acetyl, phenylacetyl) cannot be removed from the di- and tripeptide derivatives without substantial destruction of all amide bonds. These limitations are removed by using a carbobenzyloxy group in the synthesis of a β-thiolactone as illustrated in FIGURE 7. *N*-carbobenzyl-

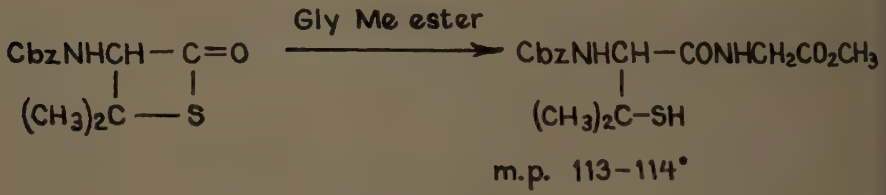


FIGURE 8. A peptide derivative from a β-thiolactone.

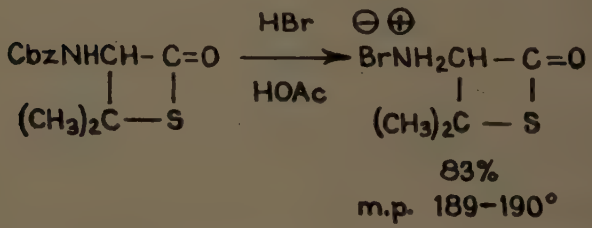


FIGURE 9. 3,3-Dimethylcysteine thiolactone hydrobromide.

oxy β,β-dimethylcysteine is cyclized in 84 per cent yield to the crystalline β-thiolactone by the use of dicyclohexylcarbodiimide. The infrared spectrem of this thiolactone shows a strong carboxyl band at 1755 to 1760 cm.$^{-1}$ No spectral data has been reported previously for β-thiolactones.

With glycine methyl ester the peptide ester is formed rapidly as illustrated in FIGURE 8.

The very interesting thiolactone hydrobromide of β,β-dimethylcysteine was obtained in crystalline form in 83 per cent yield by the hydrogen bromide cleavage of the carbobenzoxy group as illustrated in FIGURE 9. This thiolactone is analogous to homocysteine thiolactone hydroiodide obtained by Baernstein[5] and du Vigneaud.[6] However, the β-thiolactone function should be considerably more reactive than a γ-thiolactone.

References

1. WOOD, J. L. & DU VIGNEAUD. 1939. J. Biol. Chem. **130**:109.
2. SHEEHAN, J. C. & D. H. YANG. 1958. J. Amer. Chem. Soc. **80**:1158.
3. SHEEHAN, J. C., M. GOODMAN & G. HESS. 1956. J. Amer. Chem. Soc. **78**:1367.
4. KNUNYANTS, I. L. 1955. Acad. Sci. U.S.S.R. Chem. Sci. Bull. :613,619.
5. BAERNSTEIN, H. D. 1934. J. Biol. Chem. **106**:451.
6. RIEGEL, B. & V. DU VIGNEAUD. 1935-36. J. Biol. Chem. **112**:149.

CONFORMATIONAL ASPECTS OF LOW MOLECULAR WEIGHT PEPTIDES*

Murray Goodman and E. E. Schmitt
Department of Chemistry, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

In the past decade much research has been carried out on the conformation of high molecular weight polypeptides both as solids[1a-b] and in solution.[2a-g] Great strides were made, beginning with the work of Pauling and Corey[3] that showed, for example, that poly-γ-benzyl-L-glutamate can exist in the α-helical conformation.

Additional research has demonstrated that the conformation of a polypeptide chain in solution depends greatly on molecular weight,[4] solvent,[2a,5] temperature,[6] and the nature of the amino acid residue.[7] Our laboratory has been concerned with the region of chain length where helical structures can initially be formed in solution, that is, the critical size for helix formation.[8] In order to determine this type of conformational change, we undertook the step-by-step synthesis of optically pure low molecular weight peptides related to γ-methyl-L-glutamate.[9] We chose the methyl ester rather than the benzyl ester because of synthetic complications with the latter. On removal of the benzyloxycarbonyl amine blocking group, the benzyl esters were also partially removed. Our approach to the synthesis utilized mainly the active ester technique, specifically the *p*-nitrophenyl ester. The usefulness of this group has been demonstrated by Schwyzer,[10] Bodansky,[11] our laboratory,[12] and others.[13] Benzyloxycarbonyl-γ-methyl-L-glutamate (I) was prepared and allowed to react with the amino acid-active ester γ-methyl-α-*p*-nitrophenyl-L-glutamate hydrobromide (II) via the mixed anhydride method to yield the blocked dipeptide active ester (III). This compound (III) was allowed to react with amino acid (IV) and peptide esters (V) to yield the whole series of esters from the dimer through the undecamer:

1. $$\underset{\text{(I)}}{\text{Z-glu(OMe)-OH}} + \underset{\text{(II)}}{\text{Br}^{-}\,\text{H}_2^{+}\text{-glu(OMe)-O}\text{Ø}\text{NO}_2} \xrightarrow{\text{mixed anhydride}} \underset{\text{(III)}}{\text{Z-glu(OMe)-glu(OMe)-O}\text{Ø}\text{NO}_2}$$

2. $$\underset{\text{(IV)}}{\text{Z-glu(OMe)-glu(OMe)-O}\text{Ø}\text{NO}_2} + \text{H-glu(OEt)-OEt} \rightarrow \text{Z-glu(OMe)-glu(OMe)-glu(OEt)-OEt}$$

3. $$\underset{\text{(V)}}{\text{Z-glu(OMe)-glu(OMe)-O}\text{Ø}\text{NO}_2} + \text{H-glu(OMe)-[glu(OMe)]}_n\text{-glu(OEt)-OEt} \rightarrow \text{Z-glu(OMe)-[glu(OMe)]}_{n+2}\text{-glu(OEt)-OEt}$$

4. $$\underset{\text{(I)}}{\text{Z-glu(OMe)-OH}} + \underset{\text{(IV)}}{\text{H-glu(OEt)-OEt}} \xrightarrow{\text{mixed anhydride}} \text{Z-glu(OMe)-glu(OEt)-OEt}$$

* This work was supported in part by Grant G-8614 from the National Science Foundation, Washington, D. C.

where Z stands for the benzyloxycarbonyl group,

OMe
|
glu represents γ-methyl ester of glutamic acid,
$O\emptyset NO_2$ represents *p*-nitrophenyl ester, and
n is an integer from zero to nine inclusive.

Our approach allowed us to add two amino acid residues at a time to the amino end of an amino acid or peptide derivative (unpublished observations).

After an examination in the literature of the techniques used for conformational assignments, we decided to employ optical activity as the first criterion[14] for determining conformational forms of our oligomeric peptides.

TABLE 1 indicates the optical activity of the high molecular weight polymers derived from γ-benzyl-L-glutamic acid and γ-methyl-L-glutamic acid. It is of interest to note that all positive rotations in both polymers are attributable to

TABLE 1

SPECIFIC ROTATIONS OF POLY γ-BENZYL-L-GLUTAMATES AND POLY γ-METHYL-L-GLUTAMATES

Polymer	[α]	Solvent	Reference
γ-benzyl-L-glutamate	+33	*m*-cresol	15
	+14	chloroform	15, 16
	+9	dioxane	15
	+4	dimethylformamide	15
	−18	dichloroacetic acid	15
γ-methyl-L-glutamate	+65	*m*-cresol	*
	+10	dimethylformamide	*
	+100†	dioxane	*
	−35	dichloroacetic acid	16*

* Unpublished observations.
† Complicated by association since the polymer was rather low in molecular weight.

helical forms, while the negative rotations that are evident in both cases in dichloroacetic acid result from the random coil. Thus it seemed reasonable that by a measurement of the optical activity of the oligomers that we prepared, a distinction could be ascertained between helical and random coil structures. A positive rotation could be attributed to helical forms, while a negative rotation could result from random coil structures. FIGURE 1 shows the specific rotation plotted against the number of residues for the oligomeric peptides as a function of two solvents. In dichloroacetic acid the peptides exhibited a negative optical rotation approaching a value of −36.5° C. (the rotation for the high polymer). In the dioxane plot, however, there is a sharp discontinuity at the pentamer. Here we feel that there may be a definite shift from the random coil in the di-, tri-, and tetrapeptide to folded structures at the pentamer and higher. It had been pointed out in other laboratories[2a,b,4] that the rotations that are attributable to the helices are concentration-independent, which implies that all hydrogen bonds are intramolecular, and therefore association would be prohibited because of the lack of free hydrogen-bonding structures. In our oligomers, however, a definite concentration-dependence at the pentamer and higher was shown. The tri- and tetramer, to the limit

of their solubility of approximately 20 per cent, showed no concentration-dependence. The pentamer on the other hand at 0.1 per cent indicated definite association. Since in these compounds the maximum number of intramolecular hydrogen bonds is still small compared to the end groups, it is conceivable that there may still be folding of the peptide chain and association of those amides that cannot form intramolecular hydrogen bonds.[9] Molecular models

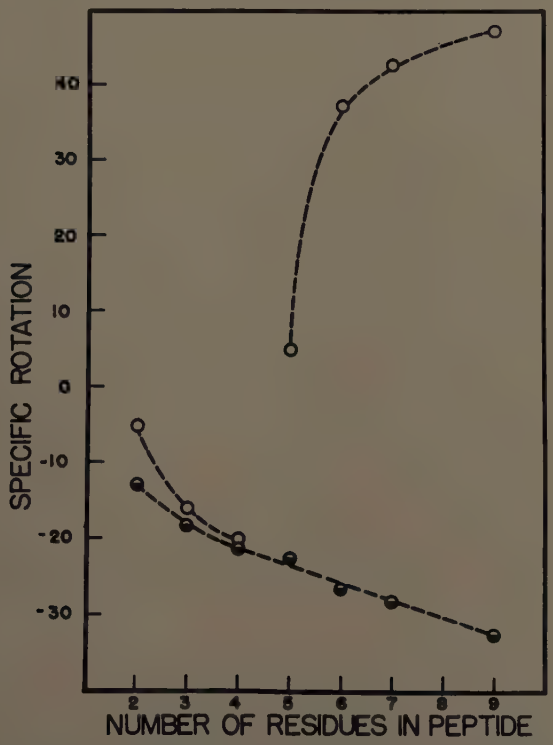


FIGURE 1. Optical activity of peptide derivatives as a function of solvent and number of residues. All rotations were measured in dioxane (open circles) and dichloroacetic acid (half-filled circles) at 2% concentration except the hepta- and nonapeptide in dioxane solution. These rotations were measured on a 1.43% and .22% solution, respectively. Reproduced by permission of the *Journal of the American Chemical Society*.[9]

indicate very definitely that there is a possibility of end group association in the oligomers that is not important for the high molecular weight analogue.

Doty and Blout[2a] synthesized a polymer mixture from the *N*-carboxy anhydride of γ-benzyl-L-glutamic acid polymerizing this to an average $\overline{DP}$ of 5.2. This polymer mixture showed very definite concentration-dependence to its optical rotation. These investigators attributed this to a transition from a β-solvated to a β-associated structure in solution as the concentration was increased. This explanation does not fit the data that we have for our compounds. We find that if the oligomers are heated in dioxane there is characteristic precipitation that results. We feel that this precipitation

actually is the β-form. If the β-form is itself not soluble in dioxane, it does not seem reasonable to include it in any explanation for the change in optical activity in solution.

Following our studies in dioxane, we attempted to find solvents where the optical activity enhancements could be found without concurrent association of the peptide chains. Robinson and Bott[17] found that copolymers of γ-methyl-L-glutamate and DL-phenylalanine gave positive optical rotations in *m*-cresol that they related to the α-form. Also, Blout and Doty[4b] found that *N*,*N*-dimethylformamide was a helix-forming solvent for poly-γ-benzyl-L-glutamate. When we investigated the optical activity of our oligomeric peptides in these two solvents, concentration-independent optical rotation enhancements were found at the nonapeptide. The question arises as to why a rather sharp discontinuity is obtained at the pentamer stage in dioxane, while in *m*-cresol and dimethylformamide (solvents where there is no association) there is a less drastic transition that is shifted to a longer chain length. Certainly the problem of conformational transitions resides in the extent of interaction between the solvent and the peptide chain. Dioxane interacts less strongly than either *m*-cresol or dimethylformamide. As a result, dimethylformamide and *m*-cresol can break up more intramolecular hydrogen bonds at the formative or critical stage such as the pentamer, hexamer, or heptamer. It is not until at least the nonamer stage in *m*-cresol and dimethylformamide that enough intramolecular hydrogen bonds can form to stabilize the structure regardless of the solvent interaction.

With the work of Moffitt,[18] Cohen,[19] Doty,[4b,6] Blout,[20] Elliott,[21] and Tinoco,[22] the use of optical rotatory dispersion to describe helical forms of polypeptides in solution has been placed on a reasonable experimental basis. Initially Moffitt and his group expanded the single-term Drude equation to include a higher ordered term with a coefficient that could be related to the per cent helix of the macromolecule.

If one plots $[\alpha]$ $(\lambda^2 - \lambda_0^2)$ against $(\lambda^2 - \lambda_0^2)^{-1}$ a straight line is obtained. If the slope is zero, Moffitt,[18] Doty,[4,15] and others[19-22] have attributed this to the random coil. Negative or positive slopes have been attributed to right- or left-handed helices. For high molecular weight polypeptides, a large amount of data has been obtained on the coefficient of the "abnormal" rotatory dispersion term b_0. For poly-γ-benzyl-L-glutamate, poly-γ-methyl-L-glutamate in helix-forming solvents, the value for b_0 of approximately -600 has been found. When our oligomeric peptides were studied in dimethylformamide and plotted according to the Moffitt equation (FIGURE 2), the results showed increasingly negative abnormal dispersions in going from the trimer to the undecamer. In dichloroacetic acid a very small positive slope was noted that was independent of chain length, indicating material essentially of a random coil nature for all the compounds. This small positive value can be viewed as the basic value for the random coil.

TABLE 2 lists the b_0 values for these peptides in dimethylformamide and dichloroacetic acid.

The situation in dioxane is complicated by molecular association. We intend to continue to investigate this system in order to explain the nature of the

association and its part in the formation of intramolecular hydrogen bonds. The data for dimethylformamide are much clearer. Since there is no association in dimethylformamide, the optical activity enhancement and abnormal rotatory dispersion indicate folding of the peptide chain. Thus we can say that for peptides derived from γ-methyl glutamate in dimethylformamide at 25° C. the critical range for helix formation has been elucidated. At the

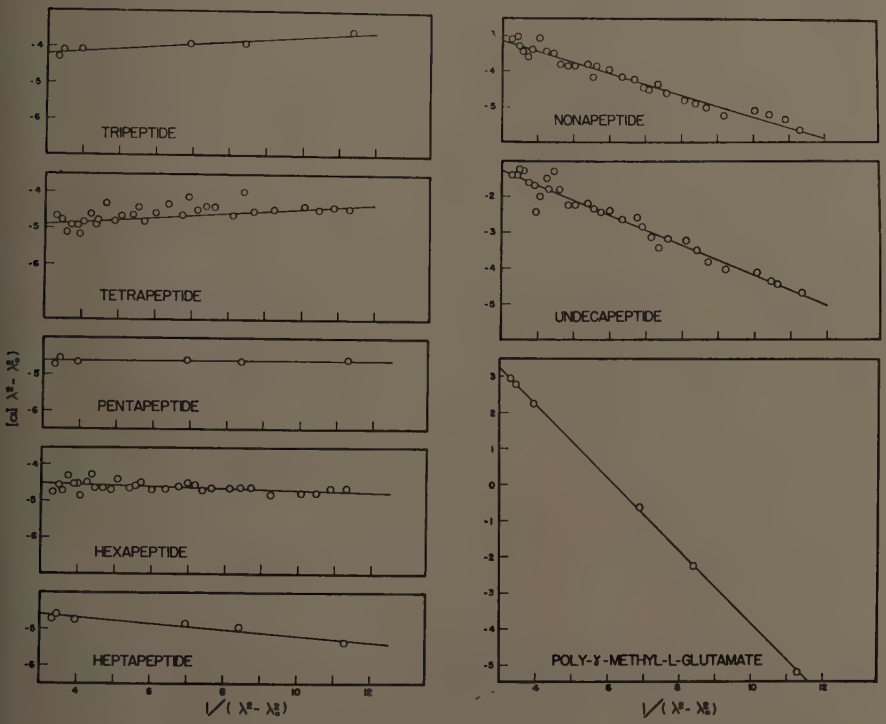


FIGURE 2. The rotatory dispersion for peptides derived from γ-methyl glutamate in dimethylformamide at 25.0° C. Reproduced by permission of the *Journal of the American Chemical Society*.[23]

TABLE 2

ABNORMAL ROTATORY COEFFICIENTS OF PEPTIDES DERIVED FROM γ-METHYL-L-GLUTAMATE*

Peptide	b_0†	Peptide	b_0†
Tri-	+52	Nona-	−191
Tetra-	+50	Undeca-	−251
Penta-	+4	Polymer DP 22	−486
Hexa-	−18	" " 93	−532
Hepta-	−53	" " high	−544

* In dichloroacetic acid a value independent of molecular weight of approximately +50 was obtained.

† In dimethylformamide solution.

nonapeptide, unmistakable helical forms are noted by both optical rotation enhancement and b_0 values. Since a small but definite change in b_0 value occurs in going from the tetra- to the pentapeptide, it is possible that this change results from the onset of folded forms in the peptide chains.

Experimental

Synthesis of oligomeric peptides. The general procedure for the preparation of the peptides consisted of allowing the blocked dipeptide-*p*-nitrophenyl ester (III) to react with a free amino group of an amino acid ester (IV) or peptide ester (V). The solvent of choice was N,N-dimethylformamide. Freshly distilled triethyl amine was used to free the amino groups from their hydrohalides. The reaction was generally allowed to react for 15 to 20 hours.

Through the hexapeptide the compounds were isolated after dilution with ethyl acetate and extraction with dilute hydrochloric acid and sodium carbonate. The peptides were recrystallized from hot ethanol. Above the hexamer the peptides were precipitated with ethanol. Trituration of the crude product with hot ethanol removed all impurities and side products.

The blocked dipeptide active ester (III) benzyloxycarbonyl-γ-methyl-L-glutamyl-γ-methyl-L-glutamic-α-*p*-nitrophenyl ester was prepared by the following general procedure: Benzyloxycarbonyl-γ-methyl-L-glutamic acid was allowed to react with the hydrobromide of γ-methyl-α-*p*-nitrophenyl-L-glutamate via the mixed anhydride using isobutylchloroformate and triethylamine.

Optical activity and rotatory dispersion measurements. All rotations were determined on a modified Rudolph Model 70 spectropolarimeter at 25° C. Corrections were made for the index of refraction of the solvents when the rotatory dispersions were measured, and a value of 212 mμ was used as the λ_0 in the Moffitt plot.

References

1*a*. Pauling, L. & R. B. Corey. 1951. Proc. Natl. Acad. Sci. U. S. **37:** 235, 241, 729.
1*b*. Pauling, L., R. B. Corey & H. R. Branson. 1951. *Ibid.* **37:** 205.
2*a*. Doty, P., A. M. Holtzer, J. H. Bradbury & E. R. Blout. 1954. J. Am. Chem. Soc. **76:** 4493.
2*b*. Blout, E. R. & M. Idelson. 1956. *Ibid.* **78:** 497.
2*c*. Berger, A. & E. Katchalski. 1951. *Ibid.* **73:** 4084.
2*d*. Katchalski, E. & P. Spitnik. 1951. *Ibid.* **73:** 2946, 3992.
2*e*. Katchalski, E. & M. Sela. 1953. *Ibid.* **75:** 5284.
2*f*. Hanby, W. E., S. G. Waley & J. Watson. 1950. J. Chem. Soc. : 3239.
2*g*. Coleman, D. 1950. *Ibid.* : 3222.
3. Pauling, L. & R. B. Corey. 1953. Proc. Roy. Soc. London. **B141:** 21.
4*a*. Doty, P., J. A. Bradbury & A. M. Holtzer. 1956. J. Am. Chem. Soc. **78:** 947.
4*b*. Mitchell, J. C., A. E. Woodward & P. Doty. 1957. *Ibid.* **79:** 3955.
5. Blout, E. R. & A. Asadourian. 1956. *Ibid.* **78:** 955.
6. Doty, P. & R. D. Lundberg. 1956. *Ibid.* **78:** 4810.
7. Bamford, C. H., A. Elliott & W. E. Hanby. 1960. Synthetic Polypeptides. : 293. Academic Press. New York, N. Y.
8. Schellman, C. G. & J. A. Schellman. 1958. Compt. rend. trav. lab. Carlsberg. Ser. chim. **30:** 465.
9. Goodman, M. & E. E. Schmitt. 1959. J. Am. Chem. Soc. **81:** 5507.
10. Iselin, B., W. Rittel, P. Sieber & R. Schwyzer. 1957. Helv. Chim. Acta. **40:** 373.
11*a*. Bodanszky, M. 1955. Nature. **175:** 685.
11*b*. Bodanszky, M. & V. du Vigneaud. 1959. J. Am. Chem. Soc. **81:** 5688.
12. Goodman, M. & K. C. Stueben. 1959. *Ibid.* **81:** 3980.
13. Wieland, T. & F. Jaenicke. 1956. Ann. Chem. Liebigs. **599:** 125.

14*a*. Doty, P. & J. T. Yang. 1956. J. Am. Chem. Soc. **78:** 498.
14*b*. Blout, E. R., P. Doty & J. T. Yang. 1957. *Ibid.* **79:** 749.
14*c*. Cohen, C. 1955. Nature. **175:** 129.
15. Yang, J. T. & P. Doty. 1957. J. Am. Chem. Soc. **79:** 761.
16. Blout, E. R., R. H. Karlson, P. Doty & B. Hargitay. 1954. *Ibid.* **76:** 4492.
17. Robinson, C. & M. J. Bott. 1951. Nature. **168:** 325.
18*a*. Moffitt, W. 1956. J. Chem. Phys. **25:** 467.
18*b*. Moffitt, W. & J. T. Yang. 1956. Proc. Natl. Acad. Sci. U. S. **42:** 596.
18*c*. Moffitt, W. 1956. *Ibid.* **42:** 736.
18*d*. Moffitt, W., D. D. Fitts & J. G. Kirkwood. 1957. *Ibid.* **43:** 723.
19. Cohen, C. & A. G. Szent-Gyorgi. 1957. J. Am. Chem. Soc. **79:** 248.
20. Blout, E. R. 1960. *In* Optical Rotatory Dispersion. Chap. 17. McGraw-Hill. New York, N. Y.
21. Bradbury, E. M., L. Brown, A. R. Downie, A. Elliott, W. E. Hanby, T. R. R. McDonald. 1959. Nature. **183:** 1736.
22. Tinoco, J. & R. W. Woody. 1960. J. Chem. Phys. **32:** 461.
23. Goodman, M., E. E. Schmitt & D. Yphantis. 1960. J. Am. Chem. Soc. In press.

NEW APPROACHES TO PEPTIDE SYNTHESIS

George W. Anderson
Lederle Laboratories, American Cyanamid Co., Pearl River, N. Y.

During the past decade, peptide synthesis has been improved by the development of new protecting groups and new reagents for forming the peptide bond. However, no development comparable in importance to Bergmann's[1a-b] "carbobenzoxy method" has emerged. The work of Bergmann and his associates in the 1930s combined the use of a new amine-protecting group, the benzyloxycarbonyl or "carbobenzoxy" group, with the azide carboxyl activating group of T. Curtius[1a-b] to make a general procedure for peptide synthesis. Today we have many useful new protecting groups and many new methods of forming the peptide bond, but definite combinations have not emerged as outstanding. I believe that such combinations will develop: for example, M. Bodanszky's nitrophenyl-ester activation with the benzyloxycarbonyl-group protection looks promising. My group at Lederle laboratories has as one of its objectives the development of improved general procedures, and this paper is concerned in large part with our efforts towards that goal.

It would be fine if the use of protecting groups could be avoided, but in practice this cannot be done. In general, the α-amino group of one amino acid or peptide and the α-carboxylic group of another must be made unreactive, so that a peptide bond can be formed between the remaining α-amino and α-carboxylic groups. Also, many amino acids have other carboxy, amino, hydroxy, or thiol groups that usually require protecting groups. An approach that appeals to us at present is to use groups that can be removed by treatment with acids for temporary protection of α-amino and α-carboxylic groups, and groups removable by catalytic hydrogenation or chemical reduction for protection of other reactive sites. In this way, the latter protecting groups can be carried through a synthesis and removed at the end. This is not a novel idea—it has been used on occasion—but perhaps its reconsideration as a general approach has value. We particularly wish to avoid the use of alkaline conditions for reasons to be given.

The increased use of the benzyl group (FIGURE 1), which can be removed by reduction and is relatively resistant to acid and base, appeals to us for protection of the imidazole-NH of histidine, the -OH of serine, threonine and tyrosine, and the -SH of cysteine. All of these are known and are being used in peptide synthesis. It should be emphasized that such protection of side groups is not always necessary, but it appears useful for a *generalized* approach.

For temporary protection of α-amino groups in the course of a peptide synthesis we have found, as have others, that Bergmann's benzyloxycarbonyl[1] has stood the test of time as a superior group. The original method of removal by catalytic hydrogenation remains valuable, and the hydrogen bromide in acetic acid method that has been developed in recent years[2,3,4] has increased the utility of the group. Thus it is possible to use benzyloxycarbonyl as one of the groups to be removed at the end of a synthesis by reduction in our general plan, or to use it for temporary protection during a synthesis.

The tert. butyloxycarbonyl group, which has been developed independently

by Carpino,[5] by Albertson and Mackay,[6] and by ourselves[7] has virtues in that it is more readily removed by acid catalysis and is insensitive to reductive procedures. The use of this group is shown in FIGURE 2. We found tert. butyl *p*-nitrophenyl carbonate to be a convenient reagent for attaching the group to amino acids, and Carpino developed the azido tert. butyl formate. At present, our experience with Carpino's reagent indicates that it is slightly superior in being easier to use and sometimes in giving better yields. A recent publication of Schwyzer *et al.*[8] confirms this opinion. An advantage of the t-BOC group (as we call it) is that gaseous by-products are formed on removal. At the present time we are exploring the possibilities of selective removal of

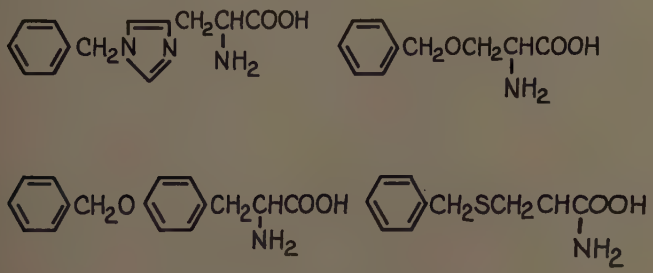


FIGURE 1. Benzyl amino acids.

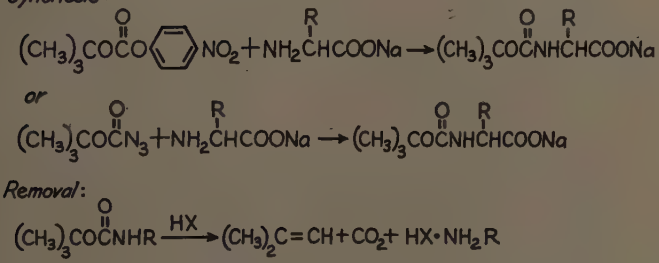


FIGURE 2. t-BOC amino acids.

this group in the presence of the benzyloxycarbonyl group, and the possibility that the group will be less sensitive than the benzyloxycarbonyl group to side reactions under alkaline conditions. An example of such a side reaction is given in FIGURE 3. Maclaren[9] found that these side reactions were serious when R_2 is H, and the reaction was carried out at room temperature with an excess of alkali.

Two amine-protecting groups resistant to mild acid treatment are the well-known tosyl and phthaloyl groups. The tosyl falls into the category of groups removable by chemical reduction; in this case, sodium in liquid ammonia is the preferred procedure, as so well exemplified by the work of Fruton[1a] and Goodman.[1b] Since this reductive procedure will also remove benzyl and benzyloxycarbonyl groups, the combined use of tosyl with these groups for protection to the end of a synthesis is useful.

The phthaloyl group has not lived up to its promised usefulness because of its sensitivity to ring opening by mild alkaline treatment and because its customary removal by hydrazine treatment presents problems when an ethyl or methyl ester is present. These esters, much used to protect carboxylic functions during a synthesis, can also react with hydrazine; perhaps more important, the liberated amino ester can undergo self-condensation reactions. Recently we have developed the use of tert. butyl esters, to be described more fully later. One of the properties of tert. butyl esters is resistance to aminolysis. Thus side reactions from hydrazinolysis of phthaloylamino esters

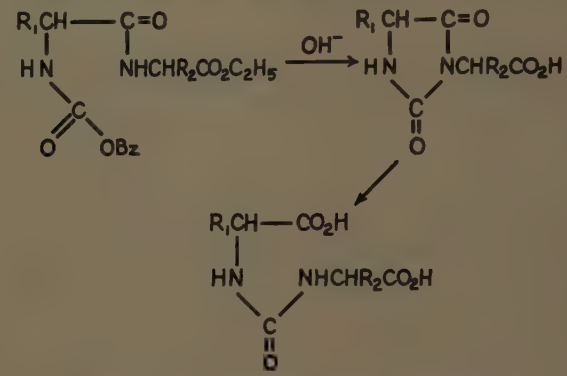


FIGURE 3. Side reactions during saponification.

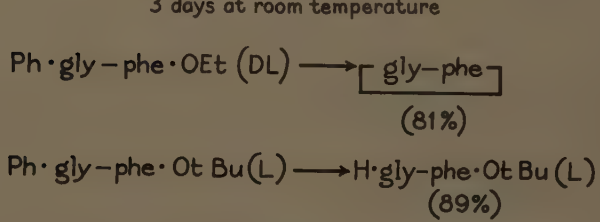


FIGURE 4. Removal of phthaloyl groups by hydrazine.

and phthaloylpeptide esters should be minimized. Our experiments to date indicate this to be true. For example, comparable reactions are shown in FIGURE 4. In the first, ethyl phthaloyl-glycyl-DL-phenylalaninate was reacted with an equivalent of hydrazine in methanol for three days at room temperature. The diketopiperazine shown precipitated with phthalhydrazide. The phthalhydrazide was removed by extraction into aqueous sodium carbonate, leaving an 81 per cent yield of the diketopiperazine, m.p. 273 to 275° C. In the second reaction shown, tert. butyl phthaloylglycyl-L-phenylalaninate was reached with hydrazine under identical conditions. The phthalhydrazide precipitated and was removed (93 per cent yield), and tert. butyl glycyl-L-phenylalaninate, m.p. 34.5 to 36° C. was obtained in 89 per cent yield by working up the filtrate. In other experiments (not shown), the phthaloyl group was removed from tert. butyl phthaloylglycinate and from tert. butyl

phthaloylglycyl-L-phenylalanyl-L-phenylalaninate to give the desired products in 72 and 62 per cent yields respectively. These experiments indicate general utility for the combined use of the phthaloyl and tert. butyl ester protecting groups. Furthermore, we have shown in another investigation[10] that tert. butyl esters can be removed by acid catalysis without affecting the phthaloyl group.

The use of tert. butyl esters for the protection of carboxylic functions is analogous to the use of the tert. butyloxycarbonyl amine-protecting group in that the ester group is readily removed by acid catalysis. Although ethyl and methyl esters have been removed by mild aqueous acid hydrolysis, there is danger of peptide-bond breakage and loss of protecting groups such as the benzyloxycarbonyl.

The reversible acid-catalyzed isobutylene reaction (FIGURE 5) has been convenient for the synthesis of tert. butyl esters. We have obtained best results by the reaction of a benzyloxycarbonylamino acid or peptide with a large excess of isobutylene in methylene chloride at room temperature for about 24 hours, using sulfuric acid as catalyst; Roeske[11] has recently reported satis-

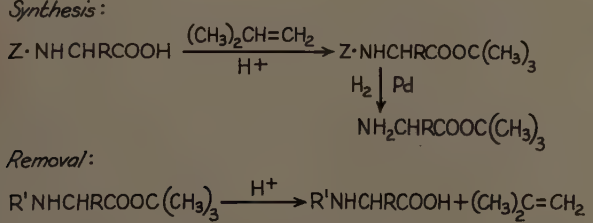


FIGURE 5. Tert. butyl esters.

factory results with several unprotected amino acids under similar conditions with dioxane solvent. The relative value of the two methods remains to be determined. In our method, the benzyloxycarbonyl group is removed by hydrogenation and the amino acid or peptide tert. butyl ester is isolated after treatment with aqueous alkali. This is used in peptide synthesis, and the tertiary butyl group is subsequently removed. When nonaqueous conditions are desired, we have found that refluxing a solution of the compound in benzene for one hour with toluene sulfonic acid is generally satisfactory for the removal.

The remarkable stability of amino acid and peptide tert. butyl esters in the form of the free bases has been indicated in the work with phthaloyl derivatives. Further evidence is given in FIGURE 6. This stability allows storage and advantageous use as the free bases.

The ordinary methyl and ethyl esters are kept as salts (usually hydrochlorides) and are frequently used in peptide synthesis as such with the addition of a base such as triethylamine. In addition to contaminating the reaction, the presence of triethylammonium chloride can lead to some racemization. We found some time ago[2] that the reaction of Z·gly-phe·OH(L) with H·gly·OET (free base) using tetraethylpyrophosphite as the reagent gave a good yield of the desired L form of the tripeptide and none of the DL form. When

H·gly·OET, hydrochloride was used with an equivalent of triethylamine, approximately 25 per cent of the racemic form was found to be present. Thus where racemization is a danger, the use of free bases is desirable. Amino acid or peptide tert. butyl esters are most convenient for this purpose.

Side reactions by alkaline hydrolysis have already been demonstrated; racemization is a problem here also. Alkaline hydrolysis can be avoided by the use of tertiary butyl esters and acid decomposition as shown in the following example.

The basic work for the comparison was an interesting investigation reported by Maclaren[9] in 1958 (FIGURE 7), who showed that the compound ethyl benzyloxycarbonylglycyl-*S*-benzyl-L-cysteinate is partially racemized during removal

Compound	% Solid formed at Room Temperature
H·gly·OEt	100 in 4 Days
H·gly·Ot Bu	25 in 325 Days
H·gly–phe·OEt(DL)	73 in 5 Hours
H·gly–phe·Ot Bu(L)	0 in 23 Days

FIGURE 6. Comparative stability of esters.

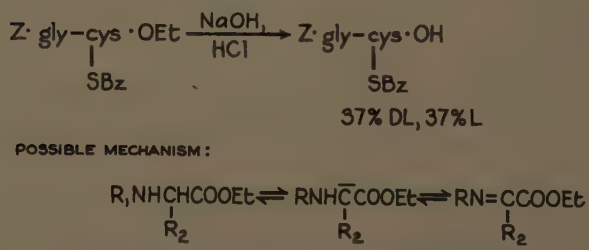


FIGURE 7. Racemization during saponification.

of the ethyl ester by alkaline saponification. The products were separated by fractional crystallization. A test showed that the desired product benzyloxycarbonylglycyl-*S*-benzyl-L-cysteine is not racemized by sodium hydroxide under the conditions of the saponification. Hence a partial deprotonization of the ester prior to the de-esterification reaction was suggested. This would not be expected under acidic conditions.

In our investigation, (FIGURE 8), H·cys·Ot·Bu was prepared by Roeske's
SBz
procedure from H·cys·OH and isolated as the phosphite salt in 55 per cent
SBz
yield. This was converted to the free base, an oil, by treatment with aqueous sodium bicarbonate solution. Reaction with Z·gly·OH by either the tetraethylpyrophosphite[2] or dicyclohexylcarbodiimide[12] procedures and recrystal-

lization of the product gave pure Z·gly-cys·Ot·Bu(L), m.p. 59.5–60° C., $\alpha]_D^{25}$ −50.3° (*c* 5, EtOH). The conversion of the ester to Z·gly-cys·OH(L) as shown in FIGURE 9 was accomplished by refluxing a benzene solution for 30 min. with toluene-sulfonic acid as catalyst, extraction into bicarbonate solution, and precipitation by dilute hydrochloric acid. After recrystallization from ethanol-water, the compound was obtained in 63 per cent yield, m.p. 124.5 to 125° C. $\alpha]_D^{25}$ −21.7° ± 2.4° (*c* 1.2, EtOH). Maclaren[9] found m.p. 120 to 121° C. $\alpha]_D^{18}$ −24° (*c* 1.2, EtOH). This reaction was deliberately carried out to less than

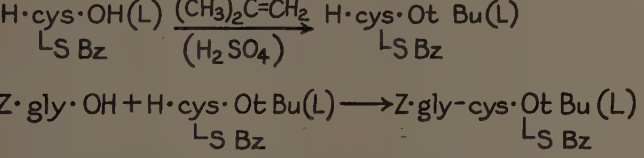


FIGURE 8. Synthesis of Z·gly-cys·Ot Bu(L).
SBz

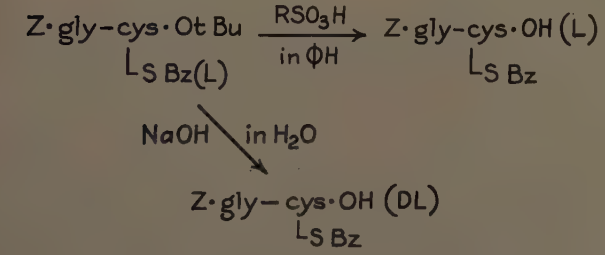


FIGURE 9. Removal of a tert. butyl ester.

completion in order to recover some of the ester; evaporation of the benzene solution and crystallization of the residue from ethyl acetate-petroleum ether gave a 16 per cent recovery of the non-racemized ester, m.p. 59 to 60° C. $\alpha]_D$ −48.7° C. (*c* 5, EtOH). Clearly the tert. butyl ester approach gives superior results as a preparative procedure. It was, then, of interest to see what the effect of alkali on the tert. butyl ester would be; it was expected that saponification would be slow under mild conditions but that the dipeptide ester would be racemized.

In a preliminary experiment in ethanol-*N* sodium hydroxide at room temperature for 40 min. the only product isolated was racemized benzyloxycarbonyl dipeptide. Our investigation is being continued to find conditions for incomplete saponification. If Maclaren's mechanism is correct, racemized ester as well as racemic acid is to be expected. It will then be interesting to compare results with other benzyloxycarbonyl dipeptide tert. butyl esters.

Before leaving the subject of protecting groups, may I mention that benzyl esters, which have had increasing use in recent years, would be useful in our

general scheme by providing protection of carboxy groups until the end of a synthesis, at which point the reductive procedure would remove them. These esters should operate well in combination with tert. butyl esters that would be used for temporary protection, since benzyl esters are more resistant to removal by acid treatment. However, experimental evidence is needed.

To summarize, the generalized use of protecting groups that we envision is:

(1) Use benzyl, tosyl, benzyloxycarbonyl (and nitro for arginine) groups for protection until the end of a synthesis. Remove them by hydrogenation or sodium in ammonia.

(2) Use tert. butyloxycarbonyl or benzyloxycarbonyl for temporary protection of amino groups during a synthesis. The trityl group[13] might also be used here. Remove groups by mild acid treatment.

(3) Use tert. butyl esters for temporary protection of carboxylic groups during a synthesis, remove by mild acid catalysis.

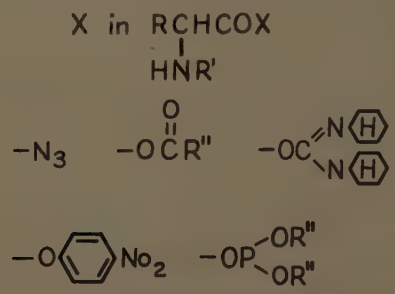


FIGURE 10. Commonly used activated intermediates.

The above procedure avoids basic conditions. It might well be possible to use phthaloyl protecting groups when tert. butyl esters are present and remove them with alcoholic hydrazine; aqueous base must be avoided.

So much for protecting groups. The other major problem in peptide synthesis is formation of the peptide bond. A number of reagents for the purpose have been reported, and many of these have not been thoroughly investigated. Today, the most commonly used methods (FIGURE 10) involve the formation of reactive derivatives of the carboxylic acid group, such as an azide, a mixed anhydride with a carbonic ester, an ester of an isourea (formed from a carbodiimide), a nitrophenyl ester, or a mixed anhydride with a phosphorus acid ester. None of the methods gives good yields without racemization in all cases; the recent work of Bodansky and du Vigneaud[14] indicates that nitrophenyl esters may have considerable promise. In any case, there is room for improved reagents; for a long peptide synthesis, yields of 90 per cent or better with no racemization are desirable. With this goal, we have sought new reagents and improvements with older reagents. Our current work with *N*,*N'*-carbonyldiimidazole and improved phosphite reagents will be discussed here.

Our interest in acyl imidazoles was stimulated by the increasing knowledge of their reactivity in biological systems and, particularly, by the evidence that the imidazole portion of histidine is a probably active site of esterases (for a

recent review, see Barnard and Stein[15]). In 1953 Wieland and Schneider[16] demonstrated that peptides could be synthesized via acylation of the imidazole ring of methyl α-*N*-benzoyl-L-histidinate, but a simpler procedure and better yields were indicated for general use. It occurred to us that *N*,*N*′-carbonyldiimidazole should be an ideal reagent. As shown in FIGURE 11, loss of carbon dioxide should be a driving force for formation of acylimidazoles. The subsequent use to form an amide, or more particularly a peptide, is also shown. While we were in the process of synthesizing *N*,*N*′-carbonyldiimidazole by the reaction of phosgene with imidazole, the synthesis of this compound by Staab appeared in the literature.[17] Using the same procedure with minor modifications (to be published), we prepared the compound and found that it does indeed react readily with carboxylic acids with evolution of carbon dioxide, and the resulting acylimidazole is reactive to amines. Our preferred procedure for peptide synthesis is to use exact equivalents of reagent and carboxylic acid in tetrahydrofuran or dimethylformamide solution. For reasons unknown, best results are obtained if the reaction solution is allowed to stand for half an hour after gas evolution ceases. In cases where racemization is not a problem

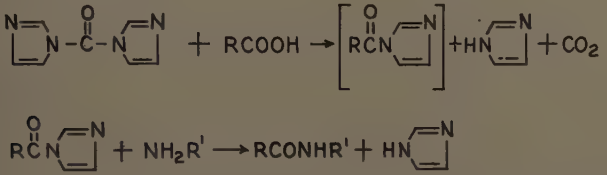


FIGURE 11. *N*,*N*′-Carbonyldiimidazole as a peptide-forming reagent.

we customarily form the acylimidazole at room temperature. In the preferred procedure, the solution is then mixed with an amino acid or peptide ester as the free base, and the peptide-forming reaction is allowed to proceed while standing overnight. Removal of the solvent under vacuum followed by washing the residue with acid, bicarbonate, and water leaves the peptide. This is then recrystallized from a suitable solvent. It has been found that amino acid or peptide ester salts can be used; the imidazole liberated in formation of the acylimidazole serves as base. However, yields were somewhat lower in several examples.

Although *N*,*N*′-carbonyldiimidazole is very reactive to water and thus acylimidazoles must be prepared in inert solvents, the acylimidazoles have much greater stability to hydrolysis. Thus it is possible to combine a solution of an acylimidazole in a water-miscible solvent with an aqueous solution of the sodium salt of an amino acid and obtain the peptide. In the few examples we have tried, yields were lower than in the nonaqueous syntheses with esters. However, it is likely that conditions can be found for improved results, and we hope to look into this.

The synthesis of several dipeptide derivatives using *N*,*N*′-carbonyldiimidazole gave good results. The next important step, which should be done with all new reagents, was a racemization study. As is well known, acyl peptides are more sensitive than acylaminoacids to racemization in peptide-forming

reactions. We have adopted as a standard in our laboratory the reaction of benzyloxycarbonylglycyl-L-phenylalanine with ethyl glycinate. In previous studies in our laboratories, we have shown that racemization can occur in this synthesis using the tetraethylpyrophosphite,[2] the mixed carbonic ester anhydride,[18] or the dicyclohexylcarbodiimide procedures,[19] and it can be eliminated in each case by the avoidance of the presence of amine hydrohalides, by change of solvent, or by the use of lower temperatures. Results found with N,N'-carbonyldiimidazole are shown in FIGURE 12. On the basis of our previous experience, the figures are accurate; the fractional crystallization procedure for separation of the racemate is quite precise. Thus again we find that racemization can be avoided by proper choice of conditions.

We are well aware of the probability of variable ease of racemization, dependent upon the actual acylpeptide that is activated. However, the accumulated experience of other workers with various reagents indicates that in many cases racemization must be small if it occurs at all. Further experience, with

$$\text{Z·gly–phe·OH(L)} + \text{H·gly·OEt} \xrightarrow{\text{CDI}} \text{Z·gly–phe–gly·OEt}$$

	Solvent	Temperature	Yield % DL	Yield % L
(1)	THF	Room	5	81
(2)*	THF	Room	8	74
(3)	DMF	−10°C.	<0.5	87

*In (2), HCl X H·gly·OEt was used

FIGURE 12. Racemization study with N,N'-carbonyldiimidazole.

careful observation, should indicate which amino acids are likely to be racemized when they occur at the carboxyl end of a peptide which is being lengthened. Then it should be possible to minimize or avoid this by proper use of reagents. In any case, we are confident that N,N'-carbonyldiimidazole belongs in the group of reagents in which racemization can be kept under control.

As a test of the general utility of N,N'-carbonyldiimidazole as a reagent, Rolf Paul of our laboratories has undertaken the synthesis of Asparagine[1] Angiotensin II (FIGURE 13). This was chosen as a model because it had already been synthesized by Rittel *et al.*[20] In addition, the naturally occurring Isoleucine[5] Angiotensin II, which differs in having an aspartyl residue in place of asparaginyl, has been synthesized by Schwarz *et al.*[21] and the valine[5] analogues of both of the above-mentioned compounds have been reported by Schwyzer *et al.*[22] We thus have an opportunity to compare our results with outstanding synthetic results by other methods. Schwarz used isobutyl chlorocarbonate and diethylchlorophosphite as reagents, and the other workers chose the azide and dicyclohexylcarbodiimide methods for most of their peptide-forming steps. Although a detailed comparison cannot be made at the present time, we are

satisfied that the over-all results will compare favorably with the other syntheses. Each of the seven peptide-forming reactions was carried out with carbonyldiimidazole reagent; in some cases, crystalline products were not obtained. The final product was purified by countercurrent distribution, and it proved to have the same order of activity in a rat assay as a sample of natural Isoleucyl Angiotensin II. At the present time, Paul is repeating the synthesis on a larger scale, and he has been able to improve yields and purity of products to the tetrapeptide stages. In the figure, Z stands for carbobenzoxy and B for tert. butyloxycarbonyl. It appears that Z derivatives crystallize more readily, but B has some advantages in very ready removal by HBr in acetic acid. Also,

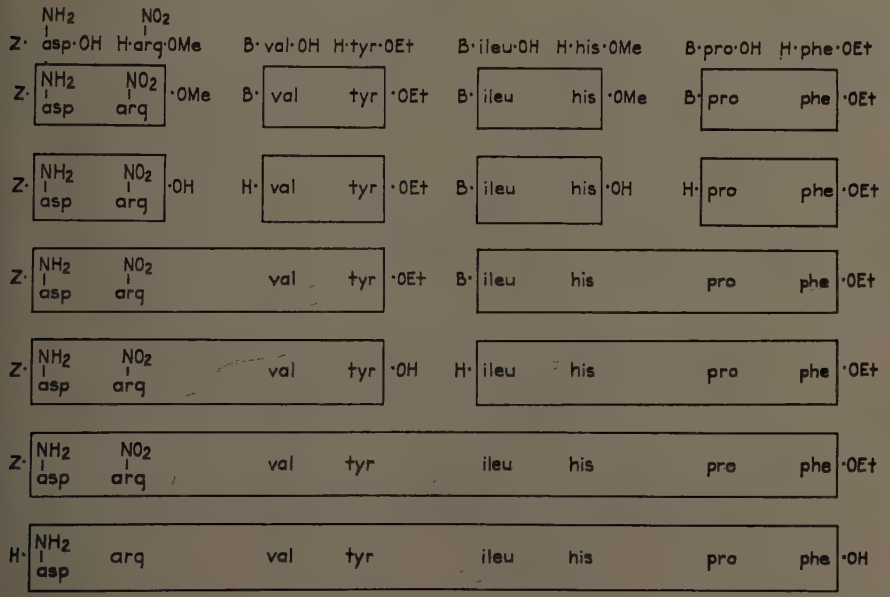


FIGURE 13. Synthesis of asparagine[1] isoleucine[5] angiotensin.

the compound B·ileu-his·OMe(L,L) is more soluble in organic solvents than the corresponding Z derivative (not shown), which was insoluble in everything, and this solubility has made its saponification less of a problem.

Another application of imidazole to peptide synthesis has been found. Since imidazole catalyzes hydrolysis of esters, it was of interest to test its effect on peptide synthesis. Preliminary experiments with mixed carbonate anhydride, azide, dicyclohexylcarbodiimide, and tetraethylpyrophosphite procedures showed a favorable effect with only the pyrophosphite. The example chosen with this reagent was the reaction of benzyloxycarbonylglycine with ethyl L-tyrosinate. Since tetraethylpyrophosphite may be used in three ways: that is, by being reacted with either peptide-forming component before the other is added ("amide" and "anhydride" procedures) or, as is generally done, by adding it to a solution of the reactants ("standard" procedure), it was of interest to discover the effect of imidazole on each of these. It was found that

the yield was decreased by the amide procedure and increased with the anhydride and standard procedures. With the latter, best results were obtained with a molar equivalent of imidazole. A fourth approach existed: the reaction of imidazole with tetraethylpyrophosphite prior to addition to the peptide-forming components. When this was tried, using an equivalent of imidazole, the best yield was obtained. The obvious conclusion is that a new, more effective reagent is formed (FIGURE 14).

A first attempt to isolate diethyl 1-imidazolephosphite by distillation was unsuccessful. However, there was evidence that the compound was present, and we expect to isolate it when we try again. Meanwhile, we were more interested in trying the second reaction in FIGURE 14, since ethylene chlorophosphite is readily available and potentially ethylene 1-imidazolephosphite would be simpler to make than tetraethylpyrophosphite itself. The reaction was performed in benzene while cooling. After removal of the precipitated triethylammonium chloride, the solvent was removed under vacuum. Recrystallization of the solid residue was accomplished from ether solution by chilling in a dry-ice acetone bath. The hygroscopic solid gave a good analysis.

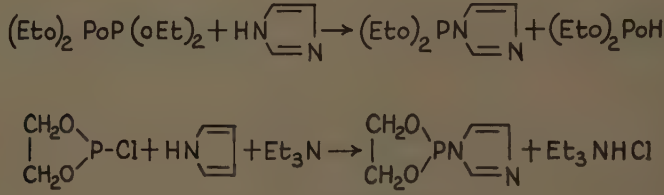


FIGURE 14. Imidazolephosphites.

When used as a reagent, the compound produced a 77 per cent yield of Z·gly-tyr·OEt(L), recrystallized. This compares with yields obtained with tetraethylpyrophosphite plus imidazole, and with yields in the 60 per cent range with tetraethylpyrophosphite alone.

Due to the hygroscopicity of the new reagent, it is somewhat difficult to handle. Consequently, we are presently looking for an analogue that is more convenient. Our thinking is that a liquid probably will be best, since it can be pipetted conveniently in the same way that we currently use tetraethylpyrophosphite.

In view of the development of new reagents, it might be expected that tetraethylpyrophosphite would fade into the background. Such is not the case in our laboratory. In our opinion, the reagent is one of the best available when properly used. It must be prepared with care; to remove all triethylammonium chloride, several distillations are necessary. The other precaution is use under strictly anhydrous conditions. In a great many cases, yields of peptides are in the 90 per cent or better bracket; in others, where yields are lower, they are generally raised by 10 per cent or more by the addition of an equivalent of imidazole.

When excellent yields can be obtained it should be possible to devise a system of carrying out several steps in a synthesis before isolating a product. Goodman and Steuben have recently published such a procedure for making tri-

peptides by combined use of dicyclohexylcarbodiimide and *p*-nitrophenylester procedures.[23] Our approach is based upon the use of tetraethylpyrophosphite as reagent and the tert. butyloxycarbonyl as the labile protecting group, as illustrated in FIGURE 15. The reaction of t-BOC phenylalanine with ethylglycinate (distilled) was accomplished by heating a diethyl phosphite solution on a steam bath for 15 min. After cooling, the solution was treated with a freshly prepared saturated solution of hydrogen bromide in diethyl phosphite to remove the t-BOC group. Reaction was immediate, and the solution was warmed on a steam bath for two minutes to complete the gas evolution. Trimethylphosphite was added as a hydrogen bromide acceptor, then benzyloxycarbonylglycine and tetraethylpyrophosphite. After 15 min. of warming, the resulting solution was cooled and the tripeptide isolated by crystallization on diluting the reaction mixture with water. The yields shown are for recrystallized, pure product. The highest yield was obtained when an equivalent of midazole was added at the beginning of the procedure.

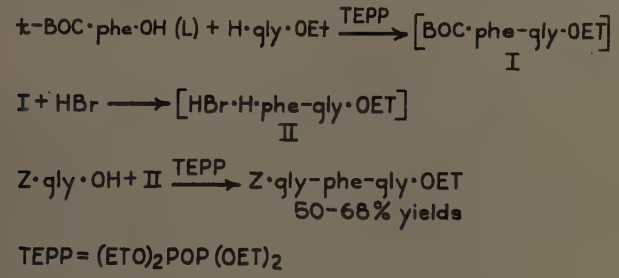


FIGURE 15. Rapid sequential procedure.

When the sequential procedure was attempted on a more complicated peptide Z·val-orn-leu-phe-pro·OMe(L,L,L,D,L) a mixture of products was obtained,

Tos

and there was evidence of incomplete reaction at the ornithine stage. The failure was not a surprise, since good yields at every step are necessary. What is needed is a way of assuring complete reaction at each stage. A simple assay procedure for each step would be helpful. We are confident that the removal of the t-BOC group is essentially quantitative, and we would not expect side reactions from hydrogen bromide in the pentapeptide used. Hence something probably happened at the peptide-forming stage, and a detection procedure for this is needed. In any case, much more work with the sequential procedure is required; perhaps its use should be restricted to tripeptides for the time being.

The peptide chemist is still looking for the synthetic procedure that will give 90 per cent or better yields at each step without racemization. Perhaps available methods could do this if optimum conditions were known. Meanwhile, it appears that the use of different reagents at various steps in the preparation of large peptides will continue to be the practice. The combinations will be chosen on the basis of the experience and prejudices of the investigator.

Sooner or later, superior general procedures will emerge. It is our belief that the process will be shortened considerably if more emphasis is placed upon fundamental studies.

References

1a. Fruton, J. 1949. Advances in Protein Chem. **5:** 1–82.
1b. Goodman, M. & G. W. Kenner. 1957. *Ibid.* **12:** 465–638.
2. Anderson, G. W., J. Blodinger & A. D. Welcher. 1952. J. Am. Chem. Soc. **74:** 5309.
3. Ben-Ishai, D. & A. Berger. 1952. J. Org. Chem. **17:** 1564.
4. Ben-Ishai, D. 1954. J. Org. Chem. **19:** 62.
5. Carpino, L. A. 1957. J. Am. Chem. Soc. **79:** 98.
6. McKay, F. C. & N. F. Albertson. 1957. J. Am. Chem. Soc. **79:** 4686.
7. Anderson, G. W. & A. C. McGregor. 1957. J. Am. Chem. Soc. **79:** 6180.
8. Schwyzer, R., P. Sieber & H. Kappeler. 1959. Helv. Chim. Acta. **42:** 2622.
9. Maclaren, J. A. 1958. Australian J. Chem. **11:** 360.
10. Anderson, G. W. & F. M. Callahan. 1960. J. Am. Chem. Soc. In press.
11. Roeske, R. W. 1959. Chem. & Ind. **1959:** 1121.
12. Sheehan, J. & G. P. Hess. 1955. J. Am. Chem. Soc. **77:** 1067.
13. Hillman-Elies, A., G. Hillman & H. Jatzkevoitz. 1953. Z. Naturforsch. **8b:** 445.
14. Bodansky, M. & V. du Vigneaud. 1959. J. Am. Chem. Soc. **81:** 5688.
15. Barnard, E. A. & W. D. Stein. 1958. Advances in Enzymol. **20:** 51.
16. Wieland, T. & G. Schneider. 1953. Ann. Chem. Liebigs. **580:** 159.
17. Staab, H. A. 1957. Ann. Chem. Liebigs. **609:** 75.
18. Vaughan, J. R., Jr. 1952. J. Am. Chem. Soc. **74:** 6137.
19. Anderson, G. W. & F. M. Callahan. 1958. J. Am. Chem. Soc. **80:** 2902.
20. Rittel, W., B. Iselin, H. Kappeler, B. Riniker & R. Schwyzer. 1957. Helv. Chim. Acta. **40:** 614.
21. Schwarz, H., F. M. Bumpus & I. H. Page. 1957. J. Am. Chem. Soc. **79:** 5697.
22. Schwyzer, R., B. Iselin, H. Kappeler, B. Riniker, W. Rittel & H. Zuber. 1958. Helv. Chim. Acta. **41:** 1287.
23. Goodman, M. & K. C. Steuben. 1959. J. Am. Chem. Soc. **81:** 3980.

SYNTHESIS OF MELANOCYTE-STIMULATING HORMONE DERIVATIVES*

Klaus Hofmann

Biochemistry Department, University of Pittsburgh School of Medicine, Pittsburgh, Pa.

Statement of Problem

The studies that provide the subject of this report were carried out during the last six years in our laboratories and are concerned with the synthesis of polypeptide derivatives embodying within their structure the entire amino acid sequence of the melanocyte-expanding hormone α-MSH (compound I, FIGURE 1).[1]

I propose to present my opinions concerning criteria for homogeneity and methods for purification of synthetic peptides, and shall discuss synthetic aspects of the α-MSH problem. I shall review in particular the preparation of three derivatives of α-MSH (compounds II, III, and IV, FIGURE 1) possessing a high degree of melanocyte-expanding activity.

Evaluation of Homogeneity of Peptides

Studies relating structure to biological activity of complex peptides can have fundamental significance only if the compounds to be tested biologically are of a high degree of homogeneity. The biological evaluation of impure samples is likely to provide misleading results. In the past, these aspects of the problem have not received the attention they deserve. Such criteria as melting points and elemental composition, which are useful when assessing the purity of low molecular weight peptides or peptide derivatives, become less meaningful as these molecules increase in size and complexity and acquire proteinlike properties.

Many complex polypeptides exhibit no sharp melting points, decompose on heating, and are frequently obtained in a solvated state. Drastic conditions are often required to remove the solvating agent, and such treatments are likely to bring about changes of the molecules invalidating the results of elemental analysis. Also, the carbon, hydrogen, and nitrogen composition of complex peptides approaches the averages found in most proteins and thus provides limited information.

Crystallinity of middle-sized peptides has been cited as evidence for homogeneity. Although important from the standpoint of purification, crystallinity as such does not constitute an absolute criterion for purity since formation

* The work reported in this paper was supported in part by Grant A-1128 from the National Institute of Arthritis and Metabolic Diseases, Public Health Service, Bethesda, Md., the National Science Foundation, Washington, D. C., the National Cancer Society, Chicago, Ill., Eli Lilly and Company, Indianapolis, Ind., and Armour and Company, Chicago, Ill.

Except where noted otherwise the amino acids and peptides mentioned in this discussion (except glycine) are of the L-variety. In the interest of space conservation the customary L-designations for individual amino acid residues are omitted.

The following abbreviations are employed: carbobenzoxy, Cbzo; *p*-toluenesulfonyl, tos or tosyl; formyl, form; acetyl, Ac; leucine amino-peptidase, LAP; *N*,*N'*-dicyclohexylcarbodiimide, DCC; nitro-L-arginine, narg; L-glutamine, gluta; benzyl, Bz; the accepted three letter designations for amino acid residues are used.

of mixed crystals or inclusion complexes can occur. Assessment of the homogeneity of crystalline preparations by more sensitive tools such as chromatography or electrophoresis on paper frequently shows them to be heterogeneous.

Formation of sharp single spots on paper chromatograms in two or more solvent systems provided one important criterion for homogeneity of intermediates and final products in our investigations. Impure samples of complex polypeptides form broad zones rather than sharp spots when they are subjected to chromatography on paper. Invariably, products exhibiting this behavior lose their tendency for trailing following rigorous purification. The fact that the compounds that form the subject of this paper contain such amino acid residues as tyrosine, methionine, histidine, arginine, and tryptophan, whose sidegroups can be detected by specific sensitive color tests, provided the opportunity to develop paper chromatograms with various color reagents. Frequently, this type of evaluation led to detection of impurities not visible by the conventional ninhydrin procedure.

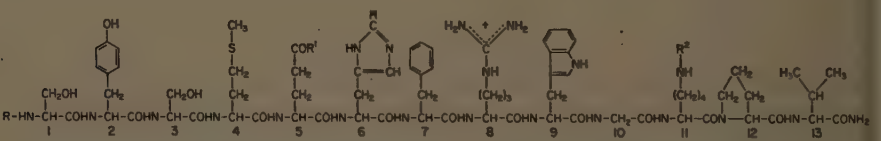


FIGURE 1. Structure of α-MSH (compound I) and three derivatives (compounds II to IV): I, R = acetyl; R^1 = OH; R^2 = H; II, R = carbobenzoxy; R^1 = NH_2; R^2 = tosyl; III, R = acetyl; R^1 = NH_2; R^2 = tosyl; IV, R = acetyl; R^1 = NH_2; R^2 = formyl.

The Partridge[2] and 2-butanol-ammonia systems[3] employed routinely in this investigation have a poor resolving power for mixtures composed of the N-carbobenzoxy derivatives of peptides. Usually these substances migrate close to the solvent front and exhibit essentially identical R_f values. Thus, formation of a single spot in these solvent systems provides no assurance for homogeneity of this type of compound.

Electrophoresis on paper was employed in some instances to check further into the purity of chromatographically homogeneous peptides; the strips were evaluated by the use of the various color reactions mentioned above. Electrophoresis was performed at various *p*H values, since it is well known that preparations which appear homogeneous at a certain hydrogen-ion concentration may separate into various components under different conditions.

Applications of these techniques to the analysis of specific peptides and peptide intermediates will be provided in a later section.

The observation that the acid hydrolysate of a synthetic peptide contains the constituent amino acids in the ratios expected by theory may provide additional evidence for homogeneity. However, this criterion is valid only if the peptide that is to be subjected to hydrolysis is pure. The intermediate products and final compounds described in this communication satisfied this criterion.

It should be noted that a hydrolysate derived from an equimolar mixture of

the precursors will exhibit the same amino acid composition as that derived from the final product.

Formation of a distribution curve of theoretical shape is another important criterion for homogeneity of peptides if conditions for establishing the curve are chosen critically. Mixtures of peptides containing such strongly basic amino acid residues as arginine and lysine are difficult to separate by countercurrent distribution (CCD) and may produce almost theoretical distribution patterns when a limited number of transfers are employed. It is only following a large number of transfers that their heterogeneity becomes apparent in many instances. Substances that exhibit apparent homogeneity on CCD may be shown to be heterogeneous when analyzed by paper chromatography.

The structure of α-MSH was determined with extremely small quantities of the hormone,[1,4] and information pertaining to the configuration of the constituent amino acids is not available; but the all-L structure seems highly probable. Based on the premise that only a molecule possessing the all-L configuration would exhibit maximal biological acitivity (an assumption that remains to be verified experimentally), my associates and I have employed only those coupling reactions that minimize the chances for racemization.

Microbiological assays of the amino acids in an acid hydrolysate were employed by Bell and his collaborators[5] for establishing the L-configuration of most of the amino acids constituting the molecule of β-corticotropin. In our own work we have used enzymatic techniques for this purpose. The well-documented, almost absolute, stereospecificity of most proteolytic enzymes, which attack only those peptide bonds in which the carbonyl group is derived from an L-amino acid residue, provided the rationale for this approach.

Leucine aminopeptidase (LAP)[6] was employed routinely, but carboxypeptidase, trypsin, and chymotrypsin were also used. Complete digestibility by LAP, with formation of the constituent amino acids in the ratios expected by theory, was accepted as evidence for stereochemical homogeneity.

Synthetic peptides have been considered to be homogeneous when their biological activity matched that of a naturally occurring counterpart. The latter must be pure for this approach to be valid. The low degree of accuracy of many biological test procedures may limit the validity of this method for assessing homogeneity.

Purification of the final product to maximal biological acitivity, without rigorous evaluation of the homogeneity of intermediates, has been used most successfully in the synthesis of such physiologically active peptides as oxytocin,[7] the vasopressins,[8,9] and hypertensin.[10]

Such an approach may not be applicable to the synthesis of more highly complex peptides since, in the course of synthesis, the introduction of a small degree of racemization into every peptide bond may result in formation of a highly complex mixture of diastereomers possessing practically identical physical and chemical properties. Such a mixture may prove to be extremely difficult to separate into individual components and, if physiological function resides only in the all-L form, will possess a low degree of biological activity. Thus careful scrutiny of the homogeneity of subunits, prior to their use in assembling the big molecules, becomes more and more critical as the complexity

of a peptide to be syntheized increases. It was this rationale that guided the work on the synthesis of the α-MSH analogues.

Purification of Peptides

Crystallization, CCD, zone electrophoresis, partition chromatography, and chromatography on ion exchangers are the methods most widely employed today for purification of polypeptides. Many of the low molecular peptides and peptide derivatives obtained in connection with our investigations were crystalline solids amenable to purification by crystallization. A number of pentapeptides and pentapeptide derivatives such as acetylseryltyrosylseryl-methionylglutamine and carbobenzoxyhistidylphenylalanylnitroarginyltrypto-phylglycine and its benzyl ester were also purified by crystallization. Homogeneous samples of the more complex peptides were prepared by CCD. Zone electrophoresis on a block of cellulose powder and CCD were used to purify the tridecapeptide derivatives (FIGURE 1, compounds II and III).

Invariably we have obtained preparations contaminated by impurities derived from the paper powder when zone electrophoresis on cellulose was employed for purification. Preparations obtained in this manner had to be subjected to CCD in order to insure a high state of purity in the final products. Chromatography on columns of carboxymethyl cellulose[11] has provided another highly promising method for purifying our complex polypeptides.

Synthetic Aspects

Stepwise elongation of a peptide chain from the carboxyl toward the amino end provides the most desirable approach to peptide synthesis at the present state of our knowledge. The exclusive use of activated acylamino acids in this method of synthesis insures formation of peptides of a high degree of stereochemical homogeneity since the activation of suitably acylated amino acids is usually not accompanied by racemization. On the other hand, racemization may ensue during activation of the C-terminal carboxyl group of acylated peptides. A synthesis of the protected nonapeptide sequence of oxytocin[12] provides an example of the application of the stepwise method for building peptide chains.

Although it is the method of choice when the synthesis of peptides in the octapeptide to decapeptide range is desired, the stepwise procedure becomes impractical when highly complex sequences are to be constructed. Deblocking creates an increasingly difficult problem as the peptide chains increase in length, and purification of complex noncrystalline peptides may be difficult and constitutes a time consuming process. Thus it becomes apparent that the combination by procedures that do not cause racemization of suitable homogeneous subunits constitutes a more practical route to highly complex synthetic polypeptides. The skillful selection of these subunits and the choice of the correct methods for their combination to produce a desired sequence is of key importance in this connection.

The peptide derivatives which are the subject of this presentation were prepared from subunits of established homogeneity in the manner shown in FIGURE 2.

The highly complex functionality of the α-MSH structure limits sharply the number of procedures that are applicable to the solution of this challenging synthetic problem. The azide,[13] mixed anhydride,[14,15,16] and *N*,*N*′-dicyclohexylcarbodiimide (DCC) procedures[17] were employed to activate carboxyl groups at selected sites. The reasons for choosing one or the other of these methods in forming certain peptide bonds will become apparent as the discussion proceeds.

The C-terminal sequence corresponding to positions 11 to 13, containing either N^{ϵ}-tosyl or N^{ϵ}-formyllysine, was linked to a suitably protected segment covering positions 6 to 10 and the ensuing blocked octapeptide amide was partially deblocked to give subunits formyl B and tosyl B, respectively. These subunits correspond to positions 6 to 13 of the α-MSH sequence. The final step in the synthesis involved the combination of the azide of the desired N-terminal pentapeptide (subunit carbobenzoxy A or acetyl A) corresponding to positions 1 to 5 with the appropriate subunit B.

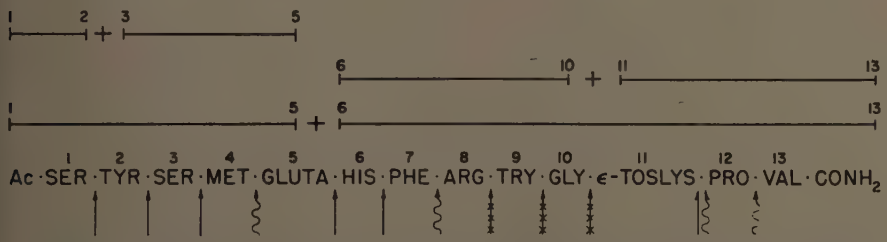


FIGURE 2. Over-all scheme used to prepare the α-MSH sequence illustrating the method employed to form each peptide bond: Arrow, azide procedure; curved arrow, mixed anhydride procedure; starred arrow, DCC procedure.

The C-termini

Subunits terminating in a glycine carboxyl group are desirable intermediates for the preparation of complex peptides since they are not expected to undergo racemization during activation. An inspection of FIGURE 1 shows that the molecule of α-MSH contains a glycine residue in position 10 and, for the reasons outlined above, the bond between this glycine and the lysine residue in position 11 was selected as a convenient site to subdivide the molecule. Thus derivatives of the C-terminal lysylprolylvaline amide sequence having a free α-amino group but containing a blocked N^{ϵ}-position were required. N^{α}-carbobenzoxy-N^{ϵ}-tosyllysylprolylvaline amide was readily obtained from N^{α}-carbobenzoxy-N^{ϵ}-tosyllysine and prolylvaline amide, and removal of the carbobenzoxy group by catalytic hydrogenation gave N^{ϵ}-tosyllysylprolylvaline amide.[18,19] This C-terminal unit was employed in preparing compounds II and III (FIGURE 1). Since attempts to remove the blocking groups from compound II with sodium in liquid ammonia resulted in fragmentation of the peptide chain, a different procedure to selectively block the N^{ϵ}-amino group of lysine was sought. A method was perfected for the preparation of N^{ϵ}-formyllysine,[20] and this compound was employed in a synthesis of N^{ϵ}-formyllysyl-

prolylvaline amide that is illustrated on FIGURE 3.* N^{ϵ}-formyllysylprolylvaline amide provided the C-terminus in the synthesis of compound IV (FIGURE 1).

The finding that the N^{ϵ}-formyl group survives the treatments that are necessary for peptide synthesis and purification, and the observation that it is readily removed from peptides under conditions that do not cause significant fission of peptide bonds (K. Hofmann and H. Yajima, unpublished observations) point to the general applicability of N^{ϵ}-formyllysine in peptide synthesis.

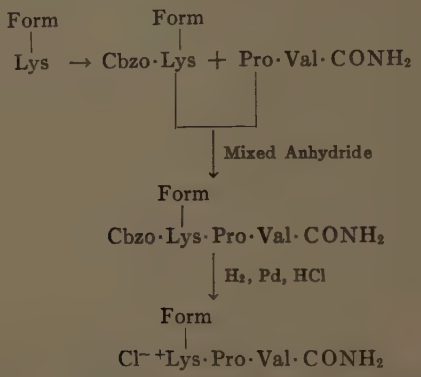


FIGURE 3. Synthetic route to partially blocked C-terminus of α-MSH (see text for details).

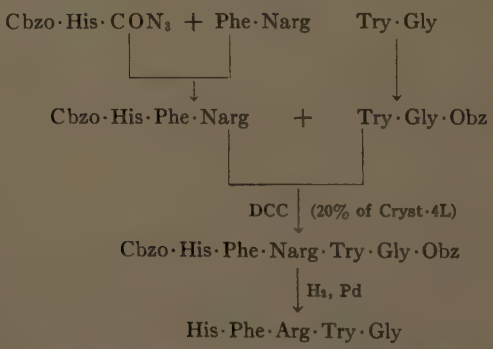


FIGURE 4. Synthetic route to central sequence (see text for details).

The Central Sequence

Carbobenzoxyhistidylphenylalanylnitroarginyltryptophylglycine was employed to attach the central sequence (positions 6 to 10) to the aforementioned C-termini. The procedure[21] for preparing this fragment, which is based on the use of nitroarginine,[22,23] is illustrated in FIGURE 4.

It was during this phase of the investigation that the application of enzymatic techniques led to detection of racemization caused by the DCC reagent.

* The synthesis of N^{ϵ}-formyl-L-ornithine; m.p. 223 to 225° C.; $[\alpha]_D^{25}$ +0.75° (in water) R_f (Partridge) = 0.30, and of N^{α}-carbobenzoxy-N^{ϵ}-formyl-L-ornithine; m.p. 116 to 118° C.; $[\alpha]_D^{27}$ −3.6° (in methanol) has also been perfected (Hofmann and Lande, unpublished observations).

Carbobenzoxyphenylalanine was coupled, by the mixed anhydride procedure, with methyl nitroarginate, and the ensuing carbobenzoxydipeptide ester was saponified and decarbobenzoxylated to give phenylalanylnitroarginine. This dipeptide in the form of its triethylammonium salt was then reacted (in the system ethyl acetate-water) with the azide of carbobenzoxyhistidine to give the crystalline carbobenzoxyhistidylphenylalanylnitroarginine. This blocked tripeptide was coupled with benzyl tryptophylglycinate (DCC serving as the condensing agent) to afford a crude carbobenzoxyhistidylphenylalanylnitroarginyltryptophylglycine benzyl ester. This material was dissolved in ethanol and, on standing, about 20 per cent of the compound separated from solution in the form of crystals; amorphous material was regenerated from the mother liquors. Both the crystalline material and the amorphous fraction were separately exhaustively hydrogenated and the ensuing hydrogenation products (histidylphenylalanylarginyltryptophylglycine) were subjected to careful ana-

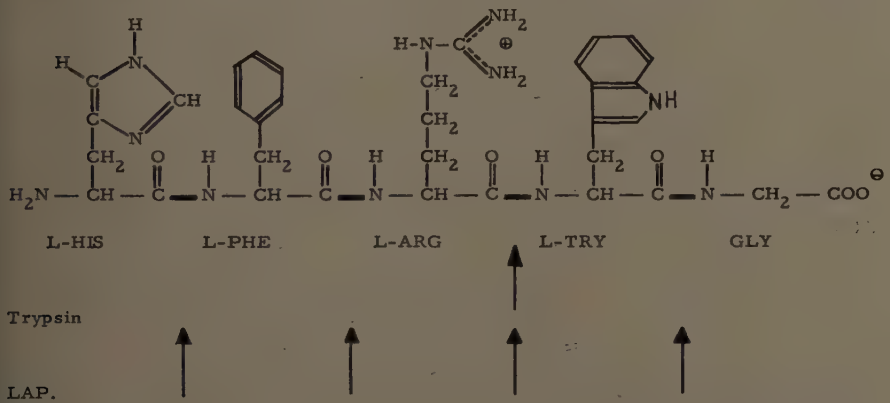


FIGURE 5. Structure of central sequence illustrating peptide bonds that are susceptible to cleavage by trypsin and LAP.

lytical evaluation. The material from both sources produced essentially one major spot on paper that reacted positively with the ninhydrin, Pauly, Sakaguchi, and Ehrlich reagents. Several minor components were seen on the chromatograms of the pentapeptide that was derived from the amorphous material. Both materials, on complete hydrolysis with acid, gave the expected amino acids (less tryptophan) in equimolar ratios, but their behavior toward proteolytic enzymes differed significantly.

An inspection of the structure of the pentapeptide (FIGURE 5) shows it to possess one trypsin labile and four LAP labile peptide linkages. Thus it was to be expected that the stereochemically homogeneous material would be converted into an equimolar mixture of histidylphenylalanylarginine and tryptophylglycine by trypsin, whereas LAP was expected to bring about fission of all four peptide bonds with liberation of the five constituent amino acids. Indeed, the material derived from the crystalline-blocked peptide exhibited exactly that type of behavior; it was completely digestible by trypsin liberating 95 per cent of the theoretical quantity of histidylphenylalanylarginine, and leucine

amino peptidase converted the compound into an equimolar mixture of the constituent amino acids. No unhydrolyzed material was present on the chromatograms from either digest. The material derived from the amorphous blocked peptide, on the other hand, gave only 37 per cent of the expected quantity of histidylphenylalanylarginine on digestion with trypsin and was only partially digestible by LAP.

Based on these results, it was concluded that the crystalline material represented the all-L form of carbobenzoxyhistidylphenylalanylnitroarginyltryptophylglycine benzyl ester and that the amorphous fraction consisted of a mixture containing the benzyl esters of carbobenzoxy-L-histidyl-L-phenylalanyl-nitro-L-arginyl-L-tryptophylglycine and of carbobenzoxy-L-histidyl-L-phenylalanyl-nitro-D-arginyl-L-tryptophylglycine. Since it could be argued that the incomplete digestibility of the pentapeptide derived from the amorphous material could have been due to the presence of impurities inhibitory to the enzymes, we have employed a strictly chemical approach to confirm the results that were obtained by the use of enzymes (K. Hofmann and S. Lande, unpublished observations). Samples of the crude pentapeptide derived from the amorphous benzyl ester were acid hydrolyzed, and arginine monohydrochloride was isolated from the hydrolysates by the flavianate procedure. This arginine monohydrochloride showed a rotation of $[\alpha]_D = \pm 1°$ in water and thus was unquestionably of the DL-variety.

A sample of L-arginine monohydrochloride that was subjected to the flavianate isolation procedure exhibited a rotation of $[\alpha]_D = +11.2°$ under identical experimental conditions.

This evidence demonstrates in an unambiguous manner that the nitroarginine moiety in carbobenzoxyhistidylphenylalanylnitroarginine had undergone marked racemization during the DCC coupling reaction with benzyl tryptophylglycinate. Two other groups of investigators[24,25] have observed racemization when using the DCC reagent in connection with other coupling reactions. Thus, the indiscriminate use of this reagent in the synthesis of complex peptide sequences is likely to cause formation of inhomogeneous materials.

Subunits B

The preparation of two sequences (subunits tosyl-B and formyl-B) that correspond to positions 6 to 13 of the α-MSH sequence will now be discussed. The scheme employed in preparing subunit tosyl-B is shown in FIGURE 6. The all-L form of carbobenzoxyhistidylphenylalanylnitroarginyltryptophylglycine (obtained in crystalline form by saponification of the pure benzyl ester)[19] was reacted with N^{ϵ}-tosyllysylprolylvaline amide in dimethylformamide with DCC serving as the condensing reagent. The ensuing mixture of compounds was exhaustively hydrogenated, and the hydrogenation products (crude subunit tosyl-B) were subjected to a 200-plate CCD. A typical pattern is shown in FIGURE 7. As may be seen from inspection of the figure, the material contained essentially three major components that corresponded respectively to peaks A, B, and C. The material corresponding to peak A was identified as histidylphenylalanylarginyltryptophylglycine, and the material corresponding to peak B was the desired histidylphenylalanylarginyltryptophylglycyl-N^{ϵ}-

tosyllysylprolylvaline amide (subunit tosyl-B). The material corresponding to the C component has not been identified unequivocally, but the available evidence points to *N*-histidylphenylalanylarginyltryptophylglycyl-*N*-cyclohexyl-*N'*-cyclohexylurea. The octapeptide amide was homogeneous when

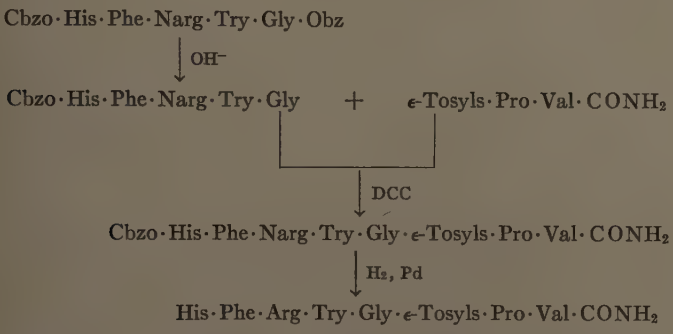


FIGURE 6. Synthetic route to subunit tosyl-B (see text for details).

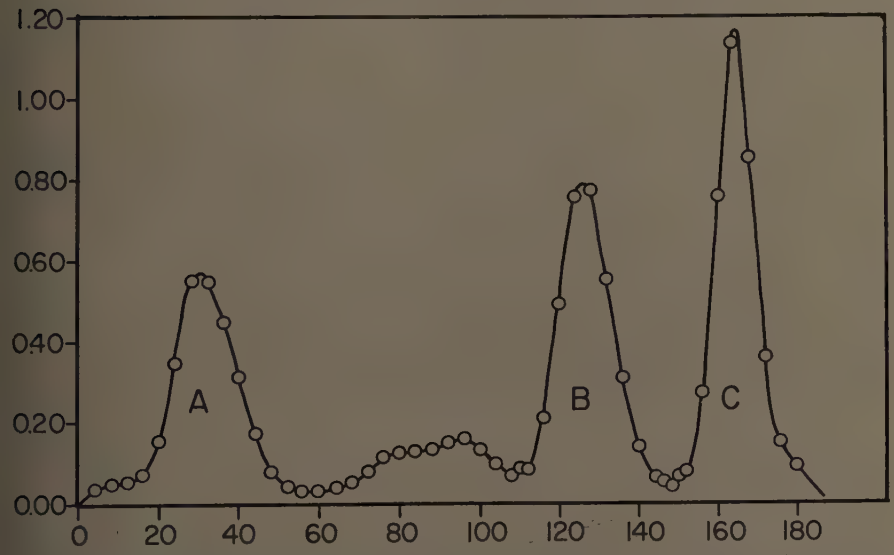


FIGURE 7. CCD pattern after 200 transfers in the solvent system *n*-butanol/5 per cent acetic acid/0.5 per cent sodium chloride of crude subunit tosyl B. Abscissa: tube number. Ordinate: absorbance at 280 mμ in lower phases.

analyzed by paper chromatography and was completely digestible by LAP with formation of an equimolar mixture of the constituent amino acids.

An identical series of reactions was employed in preparing subunit formyl-B;[20] here the N^{ϵ}-tosyl group is replaced by the formyl moiety. The mixture of compounds that resulted from the DCC coupling of carbobenzoxyhistidylphenylalanylnitroarginyltryptophylglycine with N^{ϵ}-formyllysylprolylvaline amide was hydrogenated exhaustively, and CCD separated the hydrogenation products into three components. Rather extensive distribution in two solvent

systems was required to prepare a chromatographically homogeneous sample (FIGURE 8) of histidylphenylalanylarginyltryptophylglycyl-N^{ϵ}-formyllysylprolylvaline amide (subunit formyl-B). A reproduction of a representative two-dimensional chromatogram of an LAP digest of this product is shown in FIGURE 9. The quantitative evaluation of this ninhydrin-developed chromatogram showed the presence of an equimolar mixture of the constituent amino acids.

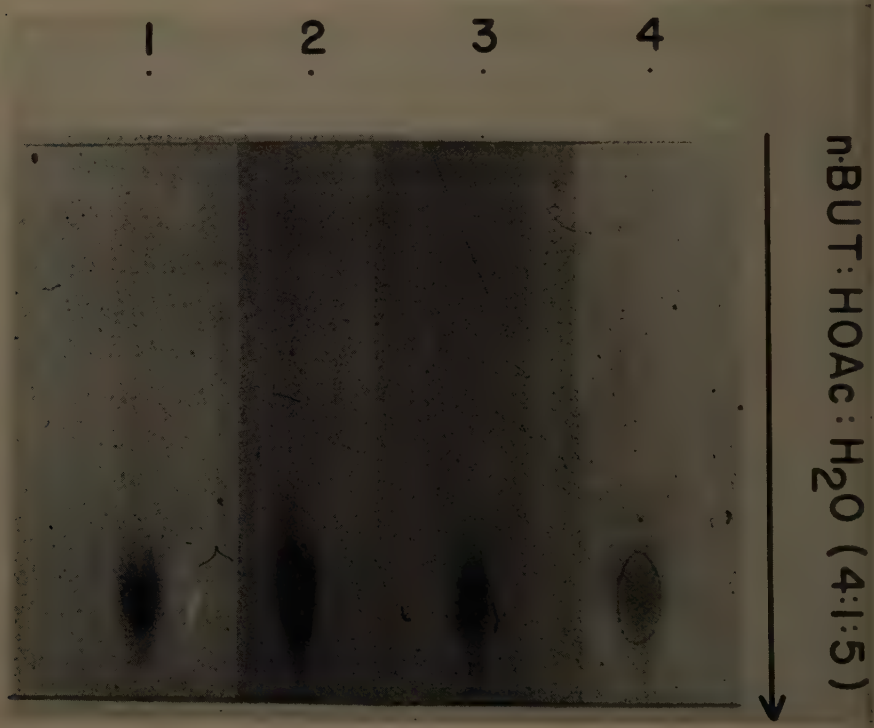


FIGURE 8. Descending unidimensional paper chromatogram of subunit formyl B in the Partridge system. Bands 1 to 4 were developed with the ninhydrin, Pauly, Ehrlich and Sakaguchi reagents, respectively.

Our experiences with the DCC reagent are not in agreement with those reported by Boissonnas and collaborators[26,27] who have employed this reagent extensively in the synthesis of peptide sequences that are closely related to those reported here. According to these investigators, the DCC reactions proceed in practically quantitative yields, and analytically pure, sharp melting reaction products are obtained without much purification.

The N-termini

In the synthesis of this portion of the α-MSH sequence the γ-carboxyl group of the glutamic acid residue was blocked by an amide function. It was expected that the replacement of glutamic acid by glutamine would not affect significantly the biological activity of the final products since biological equiv-

alence of glutamic acid and glutamine residues in peptides is not without precedence. The observation of Sondheimer and Holley[28] that the azide of carbobenzoxyglutamine couples with amino acid esters to form peptides containing glutamine was confirmed in our laboratory. We have also found[29]

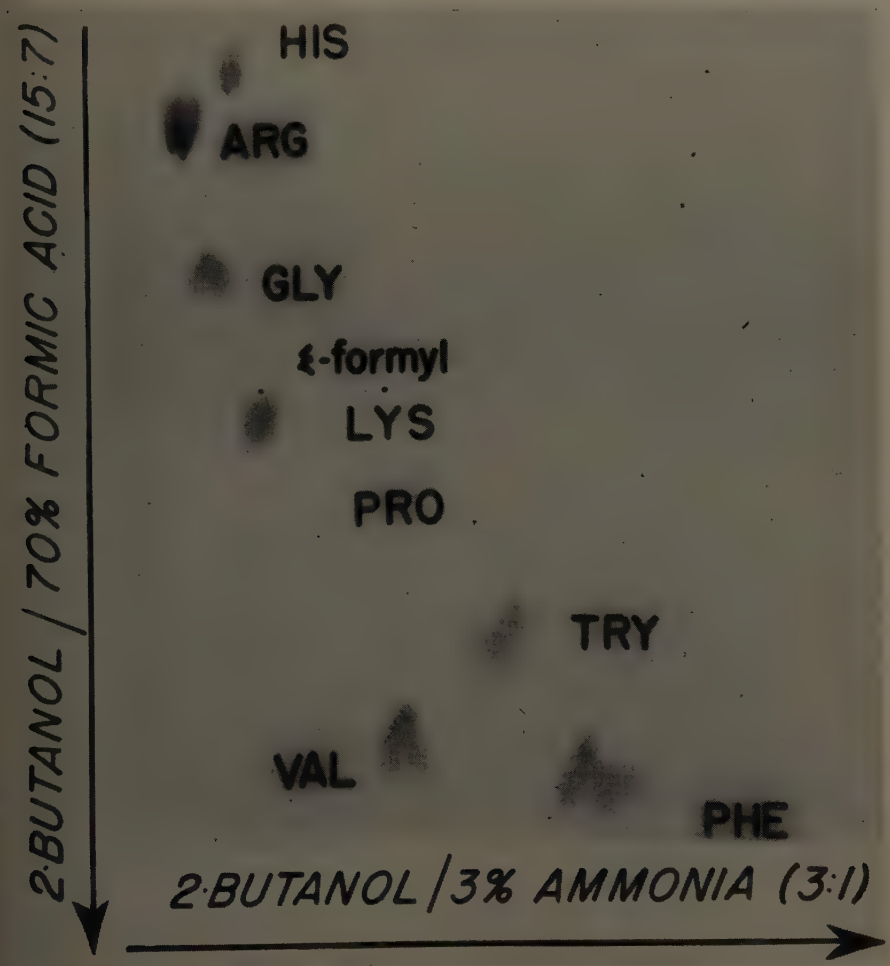


FIGURE 9. Two-dimensional paper chromatogram of LAP digest of subunit formyl B developed with the ninhydrin reagent.

that this azide reacts with simple peptide derivatives containing an N-terminal histidine to form a normal peptide bond. Complete hydrolysis of this bond by LAP excluded significant contamination with isomers involving the γ-carboxyl of glutamic acid. Based on these model experiments, the azides of carbobenzoxyseryltyrosylserylmethionylglutamine (carbobenzoxy subunit-A) and of acetylseryltyrosylserylmethionylglutamine (acetyl subunit-A) were prepared and served to introduce the N-terminus into the respective MSH analogues.

Since esters of carbobenzoxyglutamine undergo cyclization to the corresponding imides when they are subjected to the action of alkali,[30] a method[31,32] was perfected for the synthesis of peptides containing a C-terminal glutamine residue that avoided the use of esters of glutamine. This scheme, which is patterned along the procedures used in a synthesis of the N-terminal sequence of corticotropin,[33,34] is illustrated in FIGURE 10. Glutamine was acylated in an aqueous system containing triethylamine with a mixed anhydride of carbobenzoxymethionine, and the ensuing crystalline carbobenzoxymethionylglutamine was converted into methionylglutamine by sodium in liquid ammonia reduction. Carbobenzoxyseryltyrosylserylmethionylglutamine was prepared from the dipeptide by successive elongation of the chain from the amino end using a modification of the azide procedure.[34] Sodium in liquid ammonia reduction of the carbobenzoxypentapeptide gave chromatographically pure samples of seryltyrosylserylmethionylglutamine (FIGURE 11). The material was completely digestible by LAP, and the amino acid composition of the

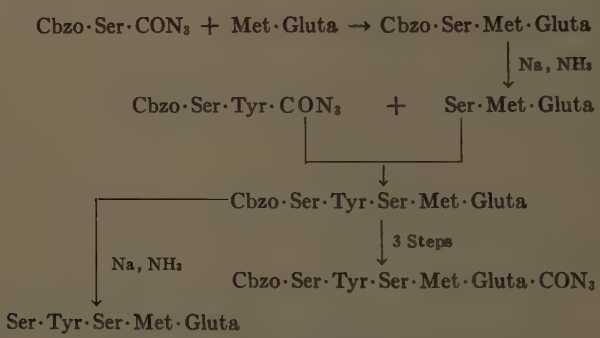


FIGURE 10. Synthetic route to subunit carbobenzoxy A (see text for details).

digest was that required by theory. The azide of the carbobenzoxypentapeptide (carbobenzoxy subunit-A) was prepared in the usual manner.

Two ninhydrin negative materials readily separable by CCD were obtained when acetic anhydride was added to a solution of seryltyrosylserylmethionylglutamine in dilute sodium bicarbonate. One of these, crystallized in needles, contained one acetyl group and exhibited a normal tyrosine ultraviolet absorption spectrum; the other, which was obtained in the form of a gelatinous mass, contained two acetyl groups and its ultraviolet absorption spectrum was markedly different. Exposure to dilute sodium hydroxide converted the gelatinous diacetyl derivative into the crystalline compound. These properties established the structure of the crystalline substance as *N*-acetylseryltyrosylserylmethionylglutamine and that of the gelatinous material as *N*-acetylseryl-O-acetyltyrosylserylmethionylglutamine. The crystalline monoacetyl derivative was converted into its hydrazide in the usual manner and the latter gave the azide (acetyl subunit-A) upon reaction with sodium nitrite. These reactions are summarized in FIGURE 12. Carboxypeptidase converted the crystalline acetylpentapeptide into an equimolar mixture of serine, tyrosine and methionine; the glutamine present on the chromatogram could not be determined quantitatively because of pyrrolidonecarboxylic acid formation and the

acetylserine escaped detection. A typical chromatogram is illustrated in FIGURE 13.

The Final Step

The three derivatives of α-MSH (FIGURE 1, Structures II, III, and IV) were obtained by combining suitable subunits A and B in the manner illustrated in FIGURE 14.[35] Thus, combination of subunit carbobenzoxy A (R = carbobenz-

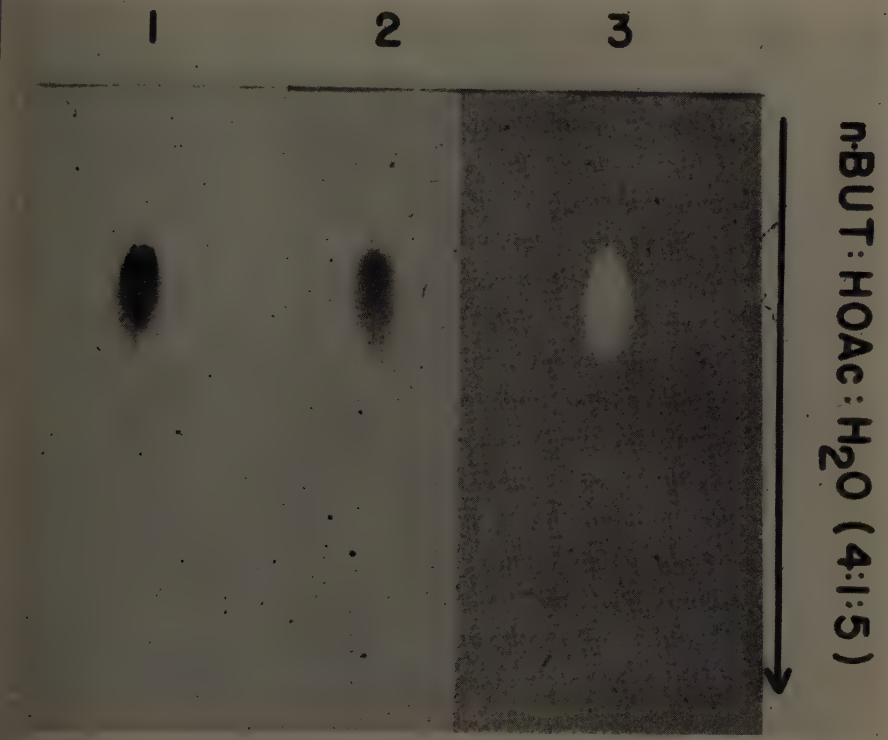


FIGURE 11. Descending unidimensional chromatogram of the pentapeptide seryltyrosylserylmethionylglutamine. Bands 1 to 3 were developed with the ninhydrin, Pauly, and chloroplatinic acid (methionine) reagents, respectively.

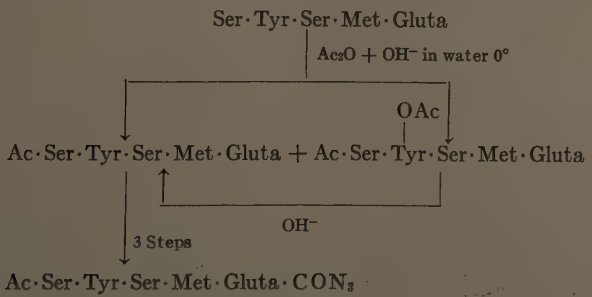


FIGURE 12. Preparation of acetyl subunit A from the free pentapeptide (see text for details).

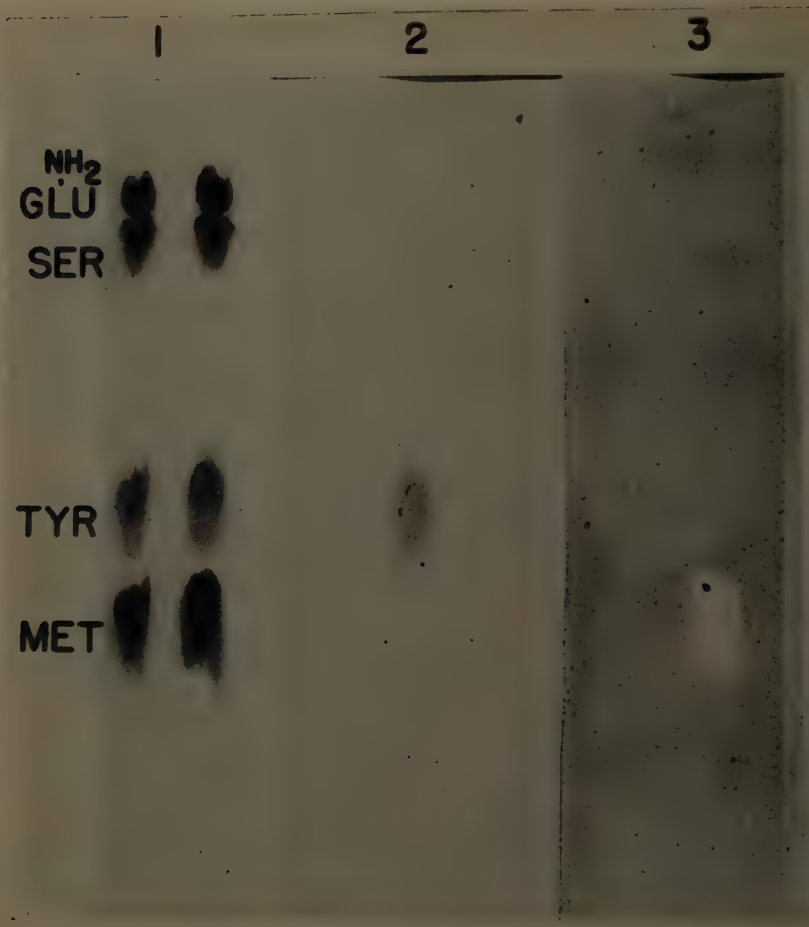


FIGURE 13. Descending unidimensional chromatogram of a carboxypeptidase digest of subunit acetyl A. Bands 1 to 3 were developed with the ninhydrin, Pauly, and chloroplatinic acid (methionine) reagents, respectively. The pattern on the extreme left is that of an equimolar mixture of the constituent amino acids. *N*-acetylserine is not detectable by these procedures.

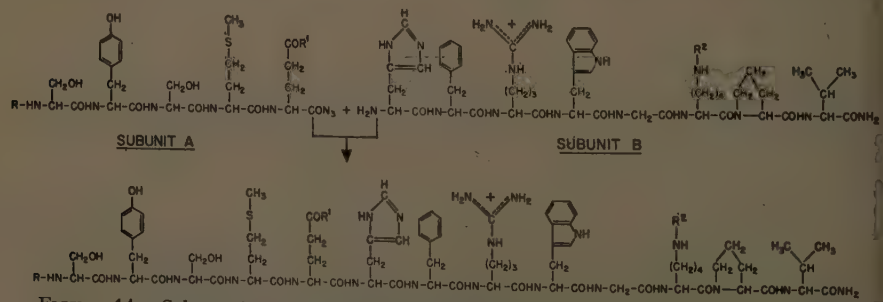


FIGURE 14. Schematic representation of final step employed in the synthesis of various derivatives of the pituitary hormone α-MSH.

oxy; $R^1 = NH_2$) with subunit tosyl B (R^2 = tosyl) gave compound II, the interaction of subunit acetyl A (R = acetyl; $R^1 = NH_2$) with subunit tosyl B (R^2 = tosyl) afforded compound III, and subunit acetyl A reacted with subunit formyl B (R^2 = formyl) to give compound IV. The reactions were carried out at a temperature of 5° by adding the appropriate solid azides to a solution

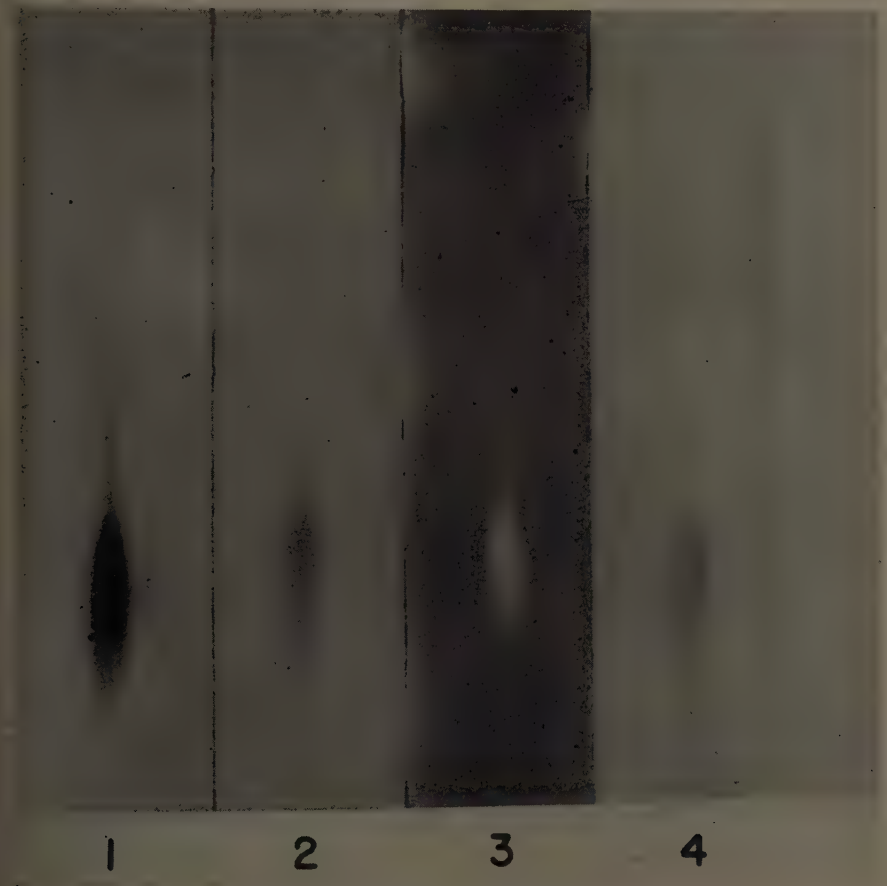


FIGURE eq. 15. Descending unidimensional chromatogram of compound IV, FIGURE 1. Bands 1 to 4 were developed with the ninhydrin, Ehrlich, chloroplatinic acid (methionine), and Sakaguchi reagents, respectively.

of the desired subunit B in dimethylformamide containing triethylamine. The decrease in the ninhydrin color of the reaction mixture was employed to assess the progress of the coupling reactions.

Extensive purification of the crude reaction products by cellulose-block electrophoresis and/or CCD was carried out. The final products, which were obtained in the form of colorless fluffy powders by freeze drying, formed sharp spots when subjected to chromatography and electrophoresis on paper, and their biological activity did not change significantly on further purification.

Acid hydrolysates of the peptides contained the constituent amino acids (except tryptophan that was destroyed) in the molar ratios predicted by theory. A representative chromatogram of compound IV and of its acid hydrolysate

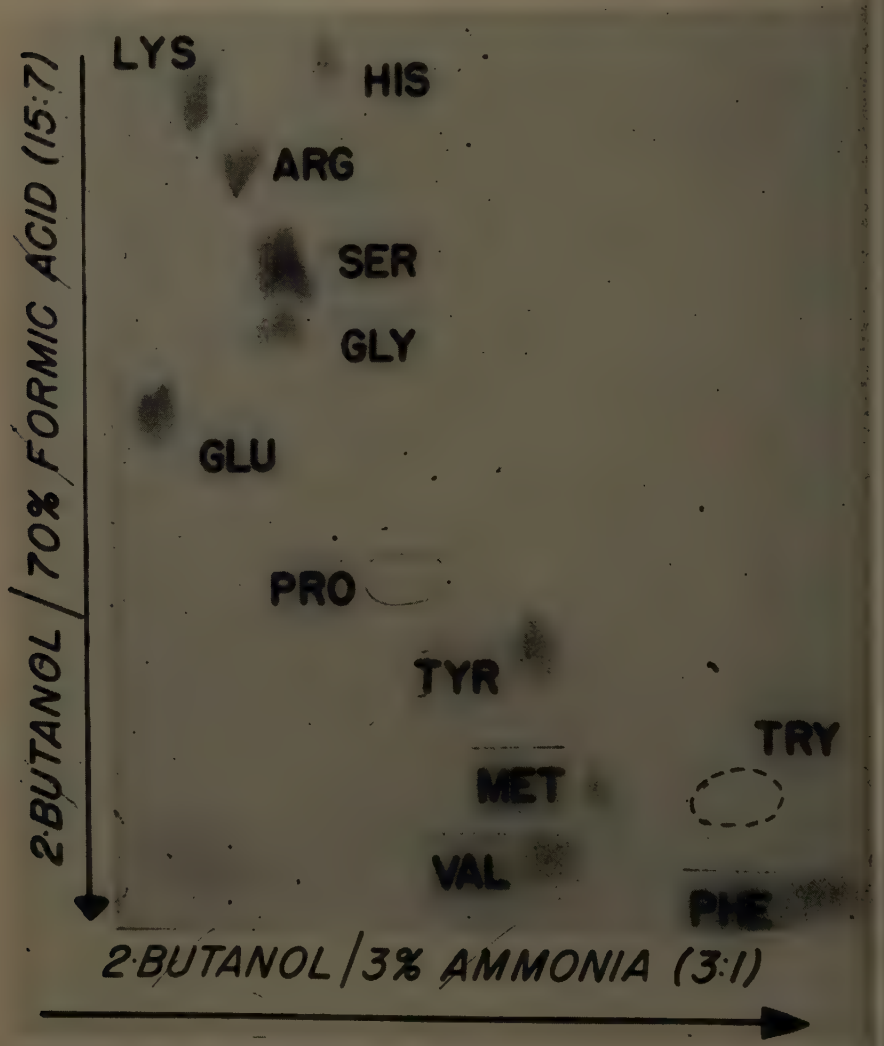


FIGURE 16. Two-dimensional paper chromatogram of an acid hydrolysate of compound IV, FIGURE 1, developed with the ninhydrin reagent.

is shown in FIGURES 15 and 16. The quantitative evaluation of the chromatogram shown in FIGURE 16 gave the amino acid ratios predicted by theory. A sample of the hydrolysate was also analyzed by the Stein-Moore technique[38] with essentially the same results; two moles of ammonia, one from the glutamine and one from the C-terminal amide group, were found to be present.

These results and the fact that the tridecapeptide derivatives were prepared from subunits of established stereochemical homogeneity by the azide method, which is known not to cause racemization, seem to justify the assignment of the 12-L configuration to these compounds.

Biological Activity and Stability

A comparison between the *in vitro* melanocyte-expanding activity of α-MSH from pig pituitary glands and that of compounds II, III, and IV whose syntheses have been described above is shown on TABLE 1.[7] Prior to discussing these results it should be pointed out that caution is indicated in the interpretation of results relating peptide structure and biological activity. This applies particularly to situations such as the one encountered with the α-MSH derivatives where the effects of blocking groups attached to a polypeptide sequence possessing high biological activity is involved. The observable biological activity may be that of the intact blocked compound or it may be a reflection of

TABLE 1
BIOLOGICAL ACTIVITY OF α-MSH AND THREE DERIVATIVES

Compound	Sequence	Biological activity/MSH units/gm.
I	Ac·Ser·Tyr·Ser·Met·Glu·His·Phe·Arg·Try·Gly·Lys·Pro·Val·$CONH_2$	1 to 2 $\times$ 10^{10}
II	Cbzo·Ser·Tyr·Ser·Met·Gluta·His·Phe·Arg·Try·Gly·Tosyls·Pro·Val·$CONH_2$	1.5 $\times$ 10^{8}
III	Ac·Ser·Tyr·Ser·Met·Gluta·His·Phe·Arg·Try·Gly·Tosyls·Pro·Val·$CONH_2$	2.0 $\times$ 10^{9}
IV	Ac·Ser·Tyr·Ser·Met·Gluta·His·Phe·Arg·Try·Gly·Formyls·Pro·Val·$CONH_2$	0.7 to 2.2 $\times$ 10^{10}

the ability of the target tissue to convert the material into the hormone. Since it is not possible, at the present stage of our knowledge, to decide between these possibilities, it is assumed in this discussion that the α-MSH derivatives possess intrinsic biological activity. It is apparent from inspection of the results presented on TABLE 1 that derivatives of the hormone in which glutamic acid is replaced by glutamine, and which contain a blocked lysine-ϵ-amino group exhibit pronounced melanocyte-expanding activity. This activity equals that of the natural hormone in the case of compound IV. These findings suggest that the presence of a free glutamic acid γ-carboxyl group and of a free ϵ-amino group of lysine is not essential for biological function.

The size of the blocking groups seems to represent a significant factor since replacement of carbobenzoxy (compound II) by acetyl (compound III) results in a marked increase in biological activity. A similar situation applies to compounds III and IV, where the exchange of tosyl (compound III) for formyl (compound IV) potentiates biological potency. We are inclined to attribute these findings to increased steric interference by the bulkier groups with the combination of these polypeptides with their receptor site.

In a recent communication Guttmann and Boissonnas[26] describe a synthesis of α-MSH, and claim that their product possessed "the physical, chemical,

and biological" properties of the natural hormone. Unfortunately these conclusions were not supported by convincing experimental evidence. First, there was no direct comparison with the natural compound, and second, the biological evaluation of the synthetic material was based on the *in vivo* Landgrebe and Waring method of assay. Since the *in vivo* activity of natural α-MSH is not known, it is impossible to ascertain the true potency of the material produced by these investigators.

Our synthetic α-MSH derivatives exhibit a high degree of stability[35] and in this respect differ markedly from the natural hormone and from the Guttmann-Boissonnas preparation; these materials are described as highly unstable, particularly when kept in the dry state.[26,37] We have stored lyophilized samples of our compounds for a span of time varying from six months to one year at a temperature of 5° C. and have observed little if any losses in biological potency on reassay. The reasons for this difference in stability are unclear at this time.

Acknowledgments

We express our sincere appreciation to W. H. Stein of the Rockefeller Institute, New York, N. Y., for amino acid determinations, and to M. R. Wright and A. B. Lerner of the Department of Medicine, Yale University School of Medicine, New Haven, Conn., for the MSH assays, which were performed by the *in vitro* method of Shizume *et al.*[38]

References

1. Harris, J. I. & A. B. Lerner. 1957. Nature. **179:** 1346.
2. Partridge, S. M. 1948. Biochem. J. **42:** 238.
3. Roland, J. F., Jr. & A. M. Gross. 1954. Anal. Chem. **26:** 502.
4. Harris, J. I. 1959. Biochem. J. **71:** 451.
5. Shepherd, R. G., K. S. Howard, P. H. Bell, A. R. Cacciola, R. G. Child, M. C. Davies, J. P. English, B. M. Finn, J. H. Meisenhelder, A. W. Moyer & J. van der Scheer. 1956. J. Am. Chem. Soc. **78:** 5051.
6. Spackman, D. H., E. L. Smith & D. M. Brown. 1955. J. Biol. Chem. **212:** 255.
7. du Vigneaud, V., C. Ressler, J. M. Swan, C. W. Roberts & G. Katsoyannis. 1954. J. Am. Chem. Soc. **76:** 3115.
8. du Vigneaud, V., M. F. Bartlett & A. Jöhl. 1957. *Ibid.* **79:** 5572.
9. du Vigneaud, V., D. T. Gish, P. G. Katsoyannis & G. P. Hess. 1958. *Ibid.* **80:** 3355.
10. Rittel, W., B. Iselin, H. Kappeler, B. Riniker & R. Schwyzer. 1957. Helv. Chim. Acta. **40:** 614.
11. Peterson, E. A. & H. A. Sober. 1956. J. Am. Chem. Soc. **78:** 751.
12. Bodanszky, M. & V. du Vigneaud. 1959. *Ibid.* **81:** 5688.
13. Curtius, T. 1902. Ber. **35:** 3226.
14. Wieland, T. & R. Sehring. 1950. Ann. **569:** 122.
15. Boissonnas, R. A. 1951. Helv. Chim. Acta. **34:** 874.
16. Vaughan, J. R., Jr. 1951. J. Am. Chem. Soc. **73:** 3547.
17. Sheehan, J. C. & G. P. Hess. 1955. J. Am. Chem. Soc. **77:** 1067.
18. Hofmann, K., M. E. Woolner, H. Yajima, G. Spühler, T. A. Thompson & E. T. Schwartz. 1958. *Ibid.* **80:** 6458.
19. Hofmann, K., T. A. Thompson, M. E. Woolner, G. Spühler, H. Yajima, J. D. Cipera & E. T. Schwartz. 1960. *Ibid.* **82.**
20. Hofmann, K., E. Stutz, G. Spühler, H. Yajima & E. T. Schwartz. 1960. *Ibid.* **82.**
21. Hofmann, K., M. E. Woolner, G. Spühler & E. T. Schwartz. 1958. J. Am. Chem. Soc. **80:** 1486.
22. Hofmann, K., A. Rheiner & W. D. Peckham. 1953. *Ibid.* **75:** 6083.
23. Hofmann, K., W. D. Peckham & A. Rheiner. 1956. *Ibid.* **78:** 238.
24. Anderson, G. W. & F. M. Callahan. 1958. J. Am. Chem. Soc. **80:** 2902.
25. Schwarz, H. & F. M. Bumpus. 1959. *Ibid.* **81:** 890.

26. Guttmann, St. & R. A. Boissonnas. 1959. Helv. Chim. Acta. **42:** 1257.
27. Boissonnas, R. A., St. Guttmann, R. L. Huguenin, P. A. Jaquenoud & E. Sandrin. 1958. *Ibid.* **41:** 1867.
28. Sondheimer, E. & R. W. Holley. 1954. J. Am. Chem. Soc. **76:** 2816.
29. Hofmann, K., H. Kappeler, A. E. Furlenmeier, M. E. Woolner, E. T. Schwartz & T. A. Thompson. 1957. *Ibid.* **79:** 1641.
30. Sondheimer, E. & R. W. Holley. 1957. *Ibid.* **79:** 3767.
31. Hofmann, K., T. A. Thompson & E. T. Schwartz. 1957. *Ibid.* **79:** 6087.
32. Hofmann, K., T. A. Thompson, H. Yajima, E. T. Schwartz & H. Inouye. 1960. *Ibid.* **82.**
33. Hofmann, K. & A. Jöhl. 1955. *Ibid.* **77:** 2914.
34. Hofmann, K., A. Jöhl, A. E. Furlenmeier & H. Kappeler. 1957. *Ibid.* **79:** 1636.
35. Hofmann, K., H. Yajima & E. T. Schwartz. 1960. *Ibid.* **82.**
36. Moore, S., D. H. Spackman & W. H. Stein. 1958. Anal. Chem. **30:** 1185.
37. Lee, T. H. & A. B. Lerner. 1956. J. Biol. Chem. **221:** 943.
38. Shizume, K., A. B. Lerner & T. B. Fitzpatrick. 1954. Endocrinology. **54:** 553.

Part IV. Biosynthesis of Proteins

PARTIAL PURIFICATION OF TRANSFER RNA*

Paul C. Zamecnik and Mary L. Stephenson
John Collins Warren Laboratories, Collis P. Huntington Memorial Hospital of Harvard University, Massachusetts General Hospital, Boston, Mass.

Nineteen years ago F. Lipmann suggested[1] that the peptide-bond forming step in protein synthesis might proceed by means of a phosphoanhydride of the carboxyl-activated amino acid. Findings of the past five years have richly borne out this prophecy. The problem of how amino acid sequence is arranged with such unerring accuracy is, however, still an unsolved one. It is patent that there is great specificity in the activating enzymes,[2,3,4,5] and it seems probable that a separate activating enzyme is present for each natural amino acid. It has also become clear that the same enzyme that activates a given amino acid is also responsible for its transfer to ester linkage on a soluble RNA molecule.[6,7]

The acceptor site for the amino acid on the cytoplasmic S-RNA molecule is a terminal adenyl group,[8,9] and immediately inside this are located either one or two cytidyl units.[10,11,12,13] The presence of this terminal nucleotide constellation is essential to the amino acid acceptor function of the S-RNA.[8,14,6,9] Among S-RNA's obtained from mammalian, yeast, and microbial origins and also in S-RNA's from the same source[15] there is variation in the ability of the RNA sample to accept amino acid. This may reflect a difference in the degree to which the -pCpCpA-end group is present on the RNA as isolated. There has been some discussion in this and other publications of the possibility of the presence of "junk RNA" in the S-RNA fraction, and the following data are related to this question. It has been shown that RNA prepared by direct phenol extraction of whole yeast cells is very similar to the S-RNA prepared by the usual homogenization and centrifugation techniques.[15] Our evidence suggests that direct phenol extraction of yeast cells results in better preservation of these end groups than does preparation of the S-RNA fraction by grinding or homogenization techniques.[15] These latter procedures expose the S-RNA fraction to attack by various nucleotidases as well as to the reversible removal of the terminal mononucleotides by the action of the enzyme(s) that attaches them to the S-RNA molecule. This degradative possibility has a bearing on the question of whether there exists in the S-RNA fraction an appreciable proportion of RNA chains incapable of serving as amino acid acceptors. Following grinding of yeast in aqueous media as illustrated in TABLE 1, the S-RNA fraction obtained has a lower amino acid and terminal AMP-accepting ability (as measured by the use of ascites cell enzymes) than does the RNA obtained

* Publication No. 992 of the Harvard University, Cambridge, Mass., Cancer Commission. The work reported in this paper was supported in part by grants from the United States Atomic Energy Commission, Washington, D.C., the American Cancer Society, New York, N.Y., and the National Cancer Institute, United States Public Health Service, Bethesda, Md.

Paul M. Doty has kindly carried out experiments that indicate that the sedimentation and diffusion constants of the dye RNA are very similar to those of the RNA. It is thus unlikely that a polymer of dye RNA actually occurs.

by direct phenol extraction of the intact cells. The loss in acceptor activity cannot be restored by addition of CTP to the ascites enzyme system. Crude yeast enzyme fractions are, however, able to restore this activity. The inference is that more than the -pCpCpA-terminal nucleotides have been removed by enzymes during the S-RNA isolation procedure. It is possible that provision of other nucleotides in addition to ATP and CTP would restore the biological activity (that is, GTP, UTP, and 5-ribosyl uridine triphosphate).

We have found[12] approximately one nucleotide to be incorporated into terminal phosphodiester linkage per 98 nucleotide residues and the total amino acid acceptor ratio to be roughly comparable to this figure.[9,16] This would correspond to a single chain molecular weight of approximately 33,000. Ultra-

TABLE 1*

COMPARISON OF LABELING OF VARIOUS TYPES OF YEAST RNA BY YEAST ENZYMES AND ASCITES ENZYMES

	Mμmoles incorporated per mg. RNA											
	C^{14}-Valine						C^{14}-ATP					
	Yeast enzyme		Ascites enzyme		Both		Yeast enzyme		Ascites enzyme		Both	
	Time in minutes											
	30	60	30	60	30	60	30	60	30	60	30	60
Soluble RNA I	1.01	1.26	0.11	0.0	0.67	0.08	—	11.55	—	0.74	7.00	0.0
Soluble RNA II	0.80	0.80	0.07	0.0	0.37	—	—	8.48	—	0.34	3.25	0.0
RNA-phenol extract of whole cell I	1.07	1.29	0.94	1.12	1.30	1.18	—	21.10	—	20.70	19.15	22.05
RNA-phenol extract of whole cell II	1.43	1.69	1.48	1.45	1.36	1.47	—	25.70	—	25.15	26.80	27.80
Ribosome RNA	0.07	0.09	0.004	0.001	0.05	0.01	—	1.76	—	0.18	0.99	0.0

* Reproduced by permission from *Biochimica et Biophysica Acta*.[15]

centrifugal data from Paul M. Doty's laboratory on our S-RNA suggest a molecular weight of about 25,000. It would thus appear that perhaps the majority of polynucleotide chains in the S-RNA fraction are capable of serving as amino acid acceptors. That not all of the S-RNA fraction consists of transfer RNA is indicated, however, by fractionation studies of Canellakis and Herbert.[17] The data in TABLE 2 illustrate the difference in base composition between the S-RNA and microsomal RNA and emphasize the distinction in minor base constitution, in agreement with the observations of other workers.[18,19,20,21] In particular the high concentration of 5-ribosyl uracil evokes speculation as to its function. This unusual base is present in sufficient concentration to occur two or three times in an S-RNA chain, assuming equal distribution in all chains. The 3′-nitrogen on this 5′-ribosyl uracil group of the S-RNA may theoretically serve as an additional attachment site for activated amino acids[22] although there is no experimental evidence for this with respect to S-RNA. Alternatively this position may be a point for attachment of the ribosyl group of a

polynucleotide chain branching off at this site. In view of the unusual and rather specific metal-binding properties of RNA,[23] this 3′-nitrogen position may also serve as a site for formation of a coordination complex with a metal.

One must bear in mind the possibility that the specificity of selection of a soluble RNA molecule for a particular amino acid may depend on more than the sequence of mononucleotide bases in the RNA (cf. Holley[24]). It is necessary to consider how the soluble RNA molecule recognizes an activating enzyme binding a particular amino acid, and here it may be that large configurational changes in individual soluble RNA molecules play a role. These changes would correspond to the secondary and tertiary structure to which Linderstrøm-Lang has called attention in discussing the properties of protein molecules.[25,26] There is reason to suspect that two separate coding areas may be present in the soluble RNA molecule: one involved in the base sequence coding of the RNA for a particular amino acid, and the second in the recognition of an activating enzyme by a specific configurational structure of a

Table 2
Nucleotide Composition of Yeast RNA Samples*

	Guanylic acid	Cytidylic acid	Uridylic acid	Uracil-5-ribosyl phosphate	Thymine ribotide	6-Methyl-aminopurine ribotide
RNA extracted from whole cells	16.1	15.9	10.2	2.2	0.32	0.19
S-RNA	15.8	15.3	10.8	2.1	0.29	0.17
Ribosomal RNA	10.7	7.7	10.1	0.3	<0.04	<0.04

Molar ratio; adenylic acid = 10.0.
* Determined after alkaline hydrolysis of the RNA. Reproduced by permission from *Biochimica et Biophysica Acta*.[15]

particular RNA molecule. These coding areas may bear an analogy to the antigen-antibody recognition problem and may offer a partial explanation for the relatively large size of the S-RNA molecule in relation to the small number of mononucleotides considered necessary for sequence coding. In connection with the sequence coding operation itself, there is growing evidence that a separate enzymatic step exists in the transfer of the aminoacyl RNA to its association with the ribonucleoprotein particle where the actual peptide chain-forming step occurs.[27,28,16,29]

It is logical to suggest that the soluble RNA molecule, or a fraction of it may, by the specificity of its sequence of purine and pyrimidine bases (and possibly by other configurational changes), designate the proper site for the activated amino acid to occupy on the ribonucleoprotein template. Efforts to separate this family of soluble RNA molecules into separate species, each coded for an individual amino acid, have therefore occupied the attention of a number of laboratories during the past two years. Schweet's laboratory[30] has employed resin columns for separation; Holley's laboratory[31] has used countercurrent distribution (CCD); Lipmann's laboratory[7] has utilized electrophoresis on resin columns; and Brown *et al.*[32] have used diazotization of resins to remove specifically the tyrosine- and histidine-containing transfer RNA.

We have exploited the periodate reaction[33,34,15] that oxidizes the 2′3′-cis-hydroxyl groups of the terminal ribose moiety of the transfer RNA molecule to the dialdehyde forms. It had previously been observed that the presence of an amino acid ester on this 3′- or 2′-hydroxyl position of the terminal ribosyl group blocked the periodate reaction.[14,6,9] It therefore appeared reasonable to hope that some large group might be attached to the aldehyde groups resulting from periodate oxidation[35] so that the transfer RNA that had no attached amino acid might be separated from that bearing a single labeled amino acid. Initially we employed 2,4-dinitrophenylhydrazine for this purpose.[15] We have found, however, that the histochemical procedure used by Ashbel and Seligman[36] for identification of tissue aldehyde groups provides an addition compound with greater promise. The details of the reaction sequence are outlined in FIGURES 1*a*, *b*, and *c* in which 2-OH, 3-naphthoic acid hydrazide readily forms a hydrazone with the aldehyde resulting from periodate oxidation of the terminal ribosyl group of the RNA. Whether a single hydrazone is formed or whether a hydrazide addition to both aldehyde groups occurs, similar to the situation in carbohydrate chemistry described by Barry and Mitchell,[37] has not been determined. The resulting naphthoic acid hydrazono-RNA is colorless in solution. Addition of tetrazotized *o*-dianisidine, freshly dissolved in water, to this naphthoic acid hydrazono-RNA results in the immediate appearance of a blue color that rapidly deepens upon addition of alcohol and is firmly bound to the RNA. Since the aminoacyl-RNA esters are unstable at alkaline *p*Hs, the reaction sequences in the dye addition are carried out at a *p*H of 5.

Details concerning the concentration of reactant, time, temperature, and *p*H optima are described elsewhere.[34] The dye RNA has a maximum absorbance of 560 mμ. It has been found important to remove the excess periodate following the periodate oxidation reaction before addition of the hydroxy-naphthoic acid hydrazide. It is also essential to remove excess hydroxy-naphthoic acid hydrazide by ethyl acetate extraction followed by repeated alcohol precipitation of the RNA before adding tetrazotized *o*-orthodianisidine to the hydrazono-RNA. If these steps are combined without removal of free hydroxy-naphthoic acid hydrazide a blue dye is formed that becomes nonspecifically absorbed to RNA regardless of whether the periodate reaction has been accomplished. At alkaline *p*Hs there is a tendency for hydroxy-naphthoic acid hydrazide to react with groups other than the aldehyde groups of the terminal ribose. This provides another reason for keeping the reaction on the weakly acid side.

After exploration of several procedures for the possible separation of dye-bound RNA from RNA, we have adopted a system composed of 2.25 *M* potassium phosphate buffer at *p*H 7.5 and *n*-propanol. The dye-bound RNA precipitates out to a large extent in the phosphate buffer alone, and the remainder precipitates at the interface of the two-phased system formed by the stepwise addition of normal propanol equilibrated with phosphate to a solution of the RNA in 2.2 *M* potassium phosphate buffer at *p*H 7.5 and at 0° C. The dye-bound RNA is insoluble in this system and may be selectively removed, whereas the RNA bearing no dye does not precipitate until *n*-propanol not previously equilibrated with the phosphate buffer has been added. Hence as

(a) Periodate Oxidation
HIO4 +
Adenine
RNA
(b) Formation of Hydrazono-RNA
Adenine
RNA
Adenine
RNA

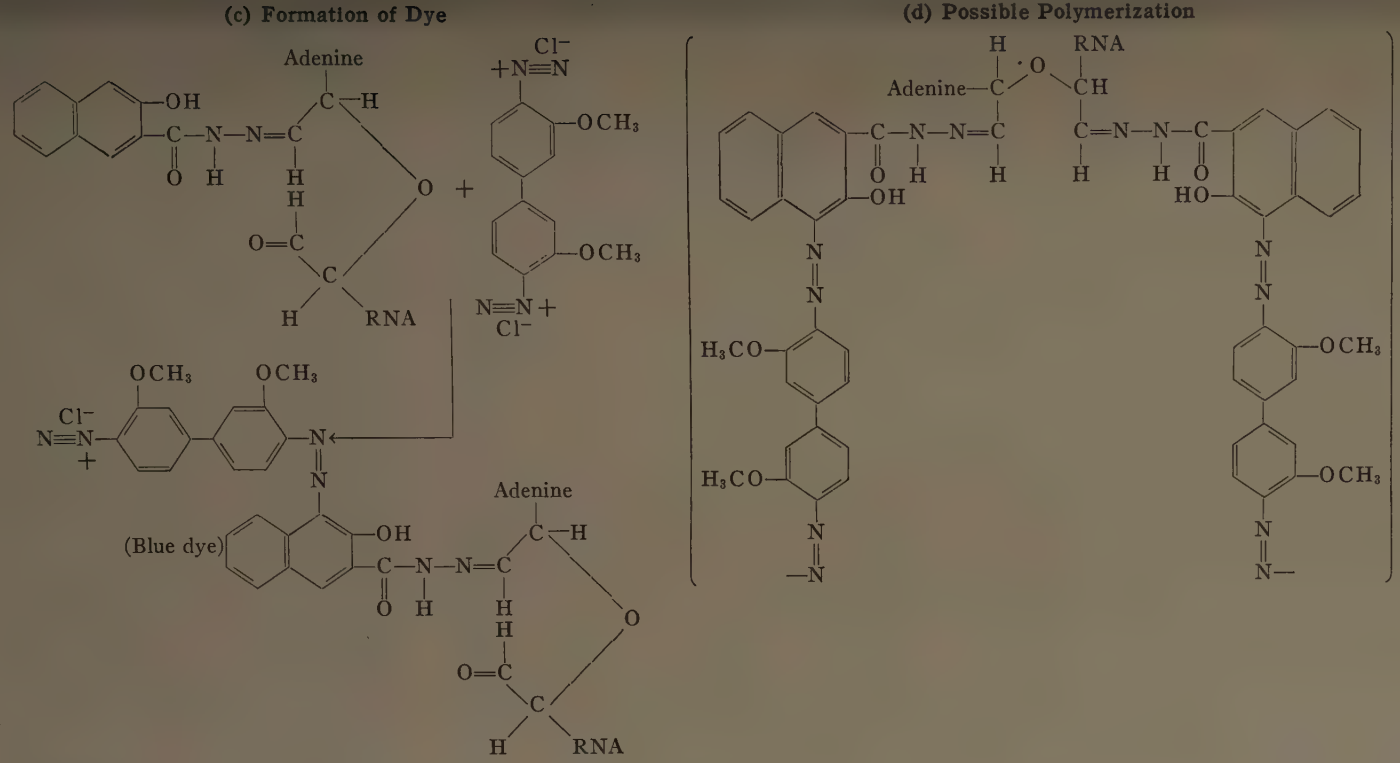


FIGURE 1. Steps in the formation of dye-RNA. Reproduced by permission from *Proceedings of the National Academy of Sciences of the United States*.[34]

water from the aqueous phase dissolves in the propanol, the phosphate buffer becomes more concentrated, and the tendency for RNA that is not dye-bound to precipitate out increases with stepwise addition of the *n*-propanol. It is not clear whether the decreased solubility of the dye-bound RNA in the phosphate-

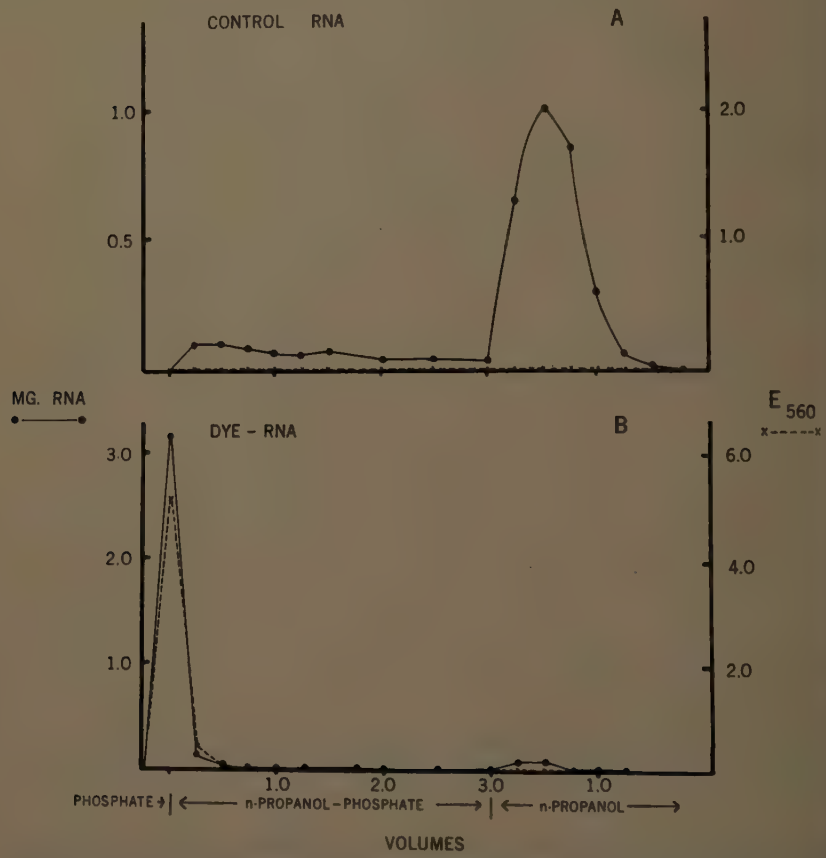


FIGURE 2. Partial fractionation of C^{14}-valine-S-RNA before and after addition of dye to the RNA. Curve A indicates the fractionation of a sample of C^{14}-valine-S-RNA. Curve B shows the fractionation of the same sample to which dye has been added. The solid line indicates the milligrams of RNA found in the interfacial precipitate and the dotted line indicates the absorption at 560 mμ. Reproduced by permission from *Proceedings of the National Academy of Sciences of the United States.*[34]

propanol system is due to the greater insolubility of the attached dye group alone or whether the disazo-coupling potentiality of the tetrazotized *o*-dianisidine has resulted in the linking together of two or more RNA molecules, a possibility suggested in FIGURE 1 (Doty, unpublished observations).

Fortunately, the optimal conditions for the production of dye-RNA are in a *p*H range in which the aminoacyl ester bond to RNA is stable. The RNA samples used in the fractionation shown in FIGURES 2 and 3 had been labeled with a single amino acid, C^{14}-valine, prior to treatment with dye. The RNA

was stripped of esterified amino acid by incubation at *p*H 10 for one hour at 37° C. A single C^{14}-labeled amino acid, L-valine, was then combined in ester linkage by means of the activation-transfer enzymatic reaction. The sample was isolated and one half was subjected to the dye-addition procedure. Both

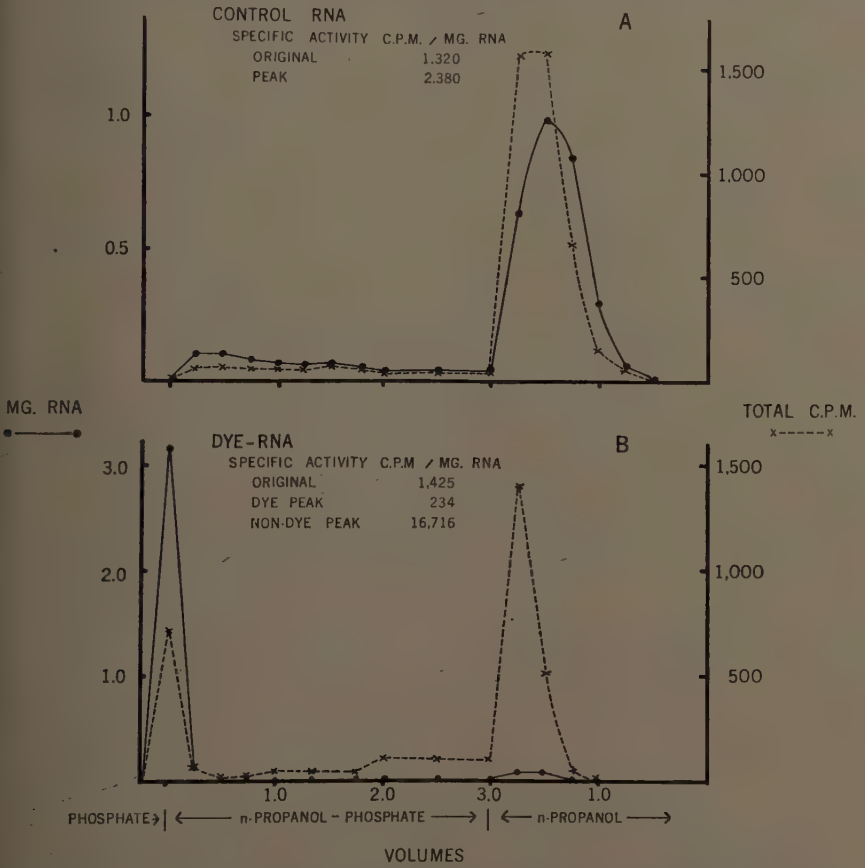


FIGURE 3. Location of radioactivity in the fractionation of C^{14}-valine-S-RNA before and after addition of dye to the RNA. The solid lines indicate the milligrams of RNA and the dotted lines show the total counts per minute found in each interfacial precipitate. The specific activity of the RNA of the unfractionated sample, the peak tube of the dye-RNA area, and the tube in the control area containing the greatest number of counts is indicated. Curve A shows the fractionation of untreated C^{14}-valine-RNA and curve B shows the fractionation of the same sample after addition of dye. Reproduced by permission from *Proceedings of the National Academy of Sciences of the United States.*[34]

samples were then fractionated by the phosphate-propanol procedure described above.

In FIGURE 2 are illustrated the results of such a fractional precipitation showing the separation of the two types of RNA. Curve A illustrates the control RNA sample and curve B the sample after treatment with dye. It may be seen in curve B that both the 260 mμ absorption peak of RNA, expressed as mg. of RNA, and the 560 mμ absorption peak of the blue dye attached to

the RNA emerge in parallel fashion, and that there is essentially no 560 absorption (that is, dye-bound RNA) in the 260 absorption peak in the same area where untreated RNA emerges (curve B).

In FIGURE 3 is illustrated the location of the C^{14}-labeled amino acid. An enhancement of specific activity of the valine-RNA of approximately 12-fold over its original specific activity was observed. Some radioactivity is associated with the dye-bound RNA fraction, presumably due to coprecipitation. It would be anticipated that an enhancement of specific activity for an individual amino acid such as valine of approximately 15- to 25-fold would occur if this were a homogeneous species of RNA coded only to acceptance of valine. This RNA fraction containing valine may therefore be expected to be heterogeneous. Direct evidence for this point of view is present in the lack of correspondence of the valine radioactivity peak with the 260 ultraviolet absorption peak of the RNA fraction. Several reasons may be offered for the heterogeneity of this fraction: first, there is the inherent difficulty in having each of the reaction steps involved in the formation of the dye-bound RNA proceed to 100 per cent completion; second, there is the instability of the single point chemical bonding of the large bulky dye molecule with the ribonucleic acid and, in this connection, we have some evidence that the dye dissociates from RNA on standing in aqueous solution over a period of hours; and, finally, it is not certain that the fraction is completely free of all amino acids except the labeled one. There is also instability of the amino acid ester-RNA bond to be considered, in explanation of why the enhancement of specific activity of the aminoacyl-RNA does not reach the theoretical value.

Bearing in mind these reservations, it appears that the above method has some promise in the quest for separation of a single species of soluble RNA coded for an individual amino acid. It would seem reasonable to expect that preliminary application of this type of fractionation procedure would also facilitate a more refined separation on the non-dye-bound RNA fraction, using a purely physical method such as that of CCD.

Acknowledgments

We express our indebtedness to Barbara Bingham for fine technical assistance, and to Jesse F. Scott, Robert B. Loftfield, and Mahlon B. Hoagland of our laboratory for suggestions and criticisms.

References

1. LIPMANN, F. 1941. Advances in Enzymol. **1:** 99.
2. HOAGLAND, M. B., E. B. KELLER & P. C. ZAMECNIK. 1956. J. Biol. Chem. **218:** 345.
3. DAVIE, E. W., V. V. KONINGSBERGER & F. LIPMANN. 1956. Arch. Biochem. Biophys. **65:** 21.
4. SYMPOSIUM ON AMINO ACID ACTIVATION. 1958. Proc. Natl. Acad. Sci. U.S. **44:** 67.
5. SCHWEET, R. S., F. C. BOVARD, E. ALLEN & E. GLASSMAN. 1958. Proc. Natl. Acad. Sci. U.S. **44:** 173.
6. PREISS, J., P. BERG, E. J. OFENGAND, F. H. BERGMANN & M. DIECKMANN. 1959. Proc. Natl. Acad. Sci. U.S. **45:** 319.
7. LIPMANN, F., W. C. HÜLSMANN, G. HARTMANN, H. G. BOMAN & G. ACS. 1959. J. Cellular Comp. Physiol. **54:** 75.
8. HECHT, L. I., M. L. STEPHENSON & P. C. ZAMECNIK. 1958. Biochim. et Biophys. Acta. **29:** 460.
9. HECHT, L. I., M. L. STEPHENSON & P. C. ZAMECNIK. 1959. Proc. Natl. Acad. Sci. U.S. **45:** 505.

10. Heidelberger, C., E. Harbers, K. C. Leibman, Y. Takagi & V. R. Potter. 1956. Biochim. et Biophys. Acta. **20:** 445.
11. Canellakis, E. S. 1957. Biochim. et Biophys. Acta. **25:** 217.
12. Hecht, L. I., P. C. Zamecnik, M. L. Stephenson & J. F. Scott. 1958. J. Biol. Chem. **233:** 954.
13. Herbert, E. 1959. Ann. N. Y. Acad. Sci. **81**(3): 679.
14. Zachau, H. G., G. Acs & F. Lipmann. 1958. Proc. Natl. Acad. Sci. U.S. **44:** 885.
15. Monier, R., M. L. Stephenson & P. C. Zamecnik. Biochim. et Biophys. Acta. In press.
16. Zamecnik, P. C. 1960. Harvey Lectures (Ser. 54, 1958–1959). : 256.
17. Canellakis, E. S. & E. Herbert. 1960. Proc. Natl. Acad. Sci. U.S. **46:** 170.
18. Dunn, D. B. 1959. Biochim. et Biophys. Acta. **34:** 286.
19. Otaka, E., Y. Hotta & S. Osawa. 1959. Biochim. et Biophys. Acta. **35:** 266.
20. Osawa, S. & E. Otaka. 1959. Biochim. et Biophys. Acta. **36:** 549.
21. Hoagland, M. B. *In* The Nucleic Acids. E. Chargaff & J. N. Davidson, Ed. Academic Press. New York, N. Y. In press.
22. Spector, L. B. & E. B. Keller. 1958. J. Biol. Chem. **232:** 185.
23. Wacker, W., E. C. & B. L. Vallee. 1959. J. Biol. Chem. **234:** 3257.
24. Holley, R. W. Ann. N. Y. Acad. Sci. In press.
25. Linderstrøm-Lang, K. U. 1952. Proteins and Enzymes. Lane Medical Lectures. Stanford Univ. Press. Stanford, Calif.
26. Doty, P., H. Boedtker, J. R. Fresco, B. D. Hall & R. Haselkorn. 1959. Ann. N. Y. Acad. Sci. **81**(3): 693.
27. Hoagland, M. B., M. L. Stephenson, J. F. Scott, L. I. Hecht & P. C. Zamecnik. 1958. J. Biol. Chem. **231:** 241.
28. Grossi, L. G. & K. Moldave. 1959. Biochim. et Biophys. Acta. **35:** 275.
29. Nathans, D. & F. Lipmann. Ann. N. Y. Acad. Sci. In press.
30. Smith, K. C., E. Cordes & R. S. Schweet. 1959. Biochim. et Biophys. Acta. **33:** 286.
31. Holley, R. W. & S. H. Merrill. 1959. J. Am. Chem. Soc. **81:** 753.
32. Brown, G. L., A. V. W. Brown & J. Gordon. 1959. Brookhaven Symposia in Biology, No. 12. Brookhaven Nat. Lab. Symposium of June 1 to 3. **1:** 47.
33. Zamecnik, P. C. & M. L. Stephenson. 1960. Federation Proc. **19:** 346.
34. Zamecnik, P. C., M. L. Stephenson & J. F. Scott. 1960. Proc. Natl. Acad. Sci. U.S. **46:** 811.
35. Whitfeld, P. R. 1954. Biochem. J. **58:** 390.
36. Ashbel, R. & A. M. Seligman. 1959. Endocrinology. **44:** 565.
37. Barry, V. C. & P. W. D. Mitchell. 1953. J. Chem. Soc. : 3610.

PURIFICATION OF A SUPERNATANT FACTOR THAT STIMULATES AMINO ACID TRANSFER FROM SOLUBLE RIBONUCLEIC ACID TO PROTEIN

Daniel Nathans*
The Rockefeller Institute, New York, N. Y.

The reactions involved in the transfer of amino acids from soluble ribonucleic acid (sRNA) to protein are in need of clarification, and it would be of obvious advantage in studying this transfer to work with as clean a system as possible, both in regard to the particle preparation and in regard to any soluble components involved. A major effort in the Lipmann laboratory† has been directed toward identifying and purifying the factors in the 105,000 *g* supernatant fraction that stimulate amino acid transfer.[1]

The initial observation by Hoagland *et al.*[2] of a stimulating effect of the *p*H-5 fraction on the transfer reaction in the rat liver microsomal system suggested that an enzyme or group of enzymes might be involved. Hülsmann and Lipmann,[3] however, observed that sulfhydryl compounds, for example glutathione or cysteine, could partially replace the 105,000 *g* supernatant in stimulating the transfer of leucine from sRNA to the protein of rat liver microsomes. In many experiments glutathione and supernatant gave nearly the same degree of stimulation. It was therefore concluded that the effect of supernatant was due, in part, to its content of glutathione, glutathione reductase, and a TPNH-generating system. The clear recognition of a sulfhydryl requirement simplified the task of recognizing other, possibly enzymic, factors involved.

Although the addition of supernatant often had little effect beyond that of sulfhydryl compounds in our microsomal system, my associates and I still suspected that additional factors were involved. If so, however, the microsomes must have contained sufficient amounts to give a near-maximal rate of transfer. To probe further into this possibility deoxycholate-extracted microsomes were tried, for Kirsch[4] at the Rockefeller Institute and Korner[5] at Cambridge University, Cambridge, England had found that deoxycholate-treated liver microsomes retained the ability to incorporate amino acids. Our procedure was as follows: rat liver microsomes were prepared in Littlefield and Keller's[6] medium A, using 24-hour fasted animals in order to deplete liver glycogen. The microsomes were well homogenized in 0.5 per cent deoxycholate and recentrifuged at 105,000 *g* for 2 hours. The pellets were rinsed several times with medium A and homogenized in the same medium. The specific activity of this preparation was about 1½ times that of the original microsomes. This preparation withstood lyophilization and storage at −20° C. for as long as 3 weeks with little or no loss of activity.

In contrast to untreated microsomes, the deoxycholate-extracted particles showed an absolute requirement for the 105,000 *g* supernatant, even in the presence of glutathione, and the deoxycholate wash was found to contain a

* Post-doctoral Research Fellow, the National Cancer Institute, Public Health Service.
† The Rockefeller Institute, New York, N. Y.

stimulatory factor. The almost complete dependence of this system on addition of supernatant is shown in FIGURE 1. This system thus could serve as a convenient assay system for supernatant factor, adjusting to values that fall

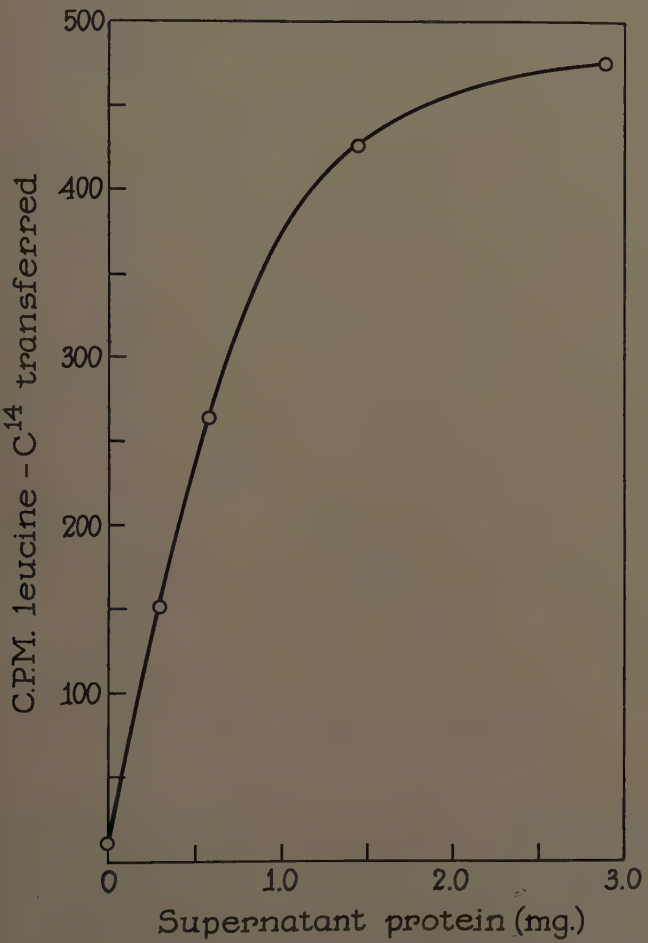


FIGURE 1. Dependence of the initial rate of leucine transfer on added 105,000 *g* supernatant. The incubation mixture contained the following in 1 ml.: deoxycholate particles (2.5 mg. protein); 0.29 mg. *Escherichia coli*-sRNA loaded with amino acids, including 1500 cpm leucine-C^{14} (10 μC/μmole); 3 μmoles of ATP; 10 μmoles of phosphoenolpyruvate; 30 μg. of PEP-kinase; 0.3 μmoles of GTP; 10 μmoles of glutathione; 100 μmoles of Tris HCl *p*H 7.0; 6 μmoles of $MgCl_2$; 50 μmoles of KCl; and supernatant as noted in the figure. After a 5 min. incubation at 35° C. 5 per cent trichloroacetic acid was added. The precipitate was extracted with 5 per cent TCA at 90° C. for 15 min., and washed with 5 per cent TCA and 3:1 alcohol-ether prior to plating for determination of radioactivity.

in the straight part of the curve. The supernatant factor is destroyed by heating at 60° C. for 5 min. and is nondialyzable. Ammonium sulfate and acetone fractionation resulted in a 30-fold purification with 30 per cent over-all recovery. The lability of the purified preparation at 0° C., however, is making further purification difficult.

FIGURE 2 shows the dependence of leucine transfer on added purified supernatant factor, and FIGURE 3 shows time curves for leucine transfer at two different concentrations of the purified fraction. In each case, 0.01 *M* glutathione was present. As noted in FIGURE 3, transfer is greatly reduced when glutathione is omitted. This glutathione effect, not apparent when whole

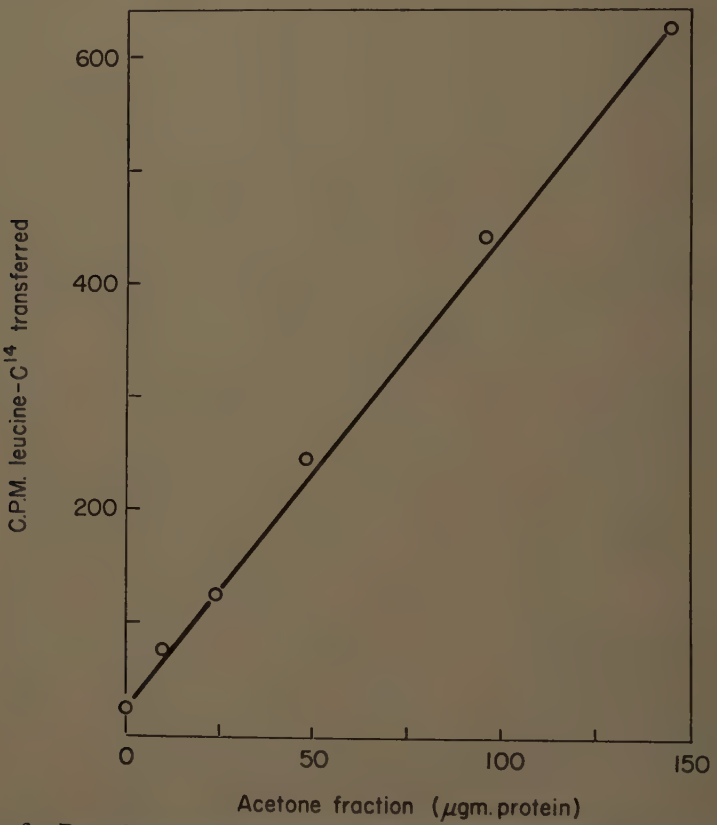


FIGURE 2. Dependence of the initial rate of leucine transfer on purified supernatant factor. The incubation mixture and conditions were the same as those given under FIGURE 1, except that 0.10 mg. of sRNA was present, and the specific activity of leucine was 25 $\mu C/\mu mole$. The purified preparation was the 26 to 32 per cent acetone fraction of the 39 to 50 per cent w./v. ammonium sulfate fraction. When this experiment was done, the purified preparation had lost one half its activity.

supernatant is used, is clearly brought out with the purified fraction. As with the untreated microsomes, cysteine is as effective as glutathione.

We conclude, therefore, that with this preparation a sulfhydryl compound and the partially purified, heat-labile, nondialyzable factor can replace the whole supernatant in stimulating leucine transfer from sRNA to the protein of microsomal particles. Although the results to date are consistent with an enzymatic action of the purified fraction, it is not yet certain that this factor is an enzyme.

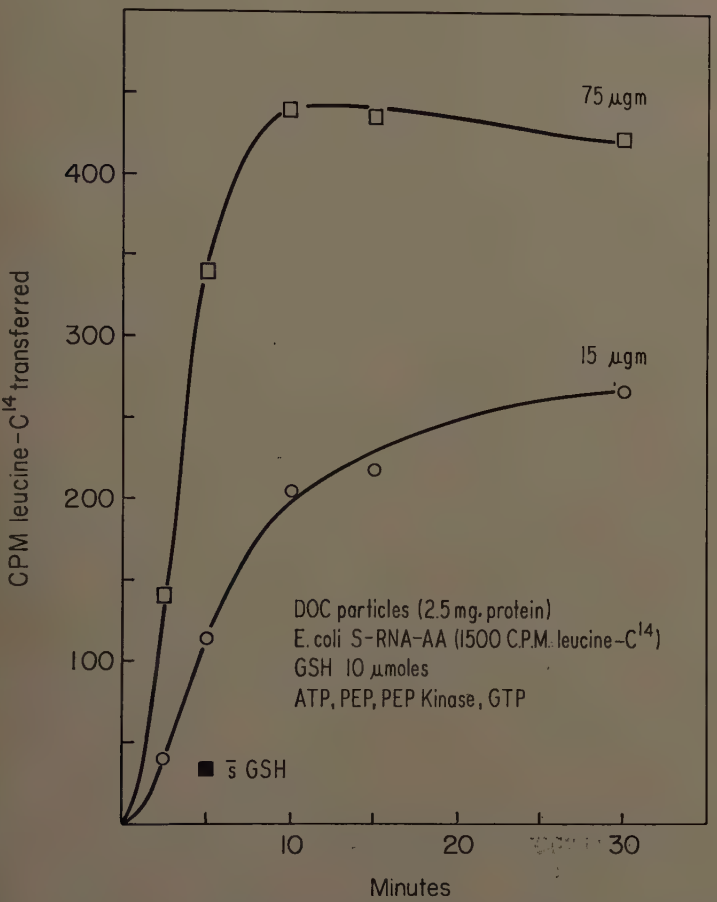


FIGURE 3. Time curves of leucine transfer with different amounts of the purified supernatant factor. The incubation mixture and conditions were the same as those given under FIGURE 1, except that 0.14 mg. of sRNA was present, and the specific activity of leucine was 25 μC/μmole. Lower curve, 15 μg. protein in the supernatant fraction; upper curve, 75 μg. The solid square indicates 5-min. value in the absence of glutathione, paired with the upper 5-min. point.

References

1. LIPMANN, F., W. C. HÜLSMANN, G. HARTMANN, H. G. BOMAN & G. ACS. 1959. J. Cellular Comp. Physiol. **54**(Supp. 1): 75.
2. HOAGLAND, M. B., M. L. STEPHENSON, J. F. SCOTT, L. I. HECHT & P. C. ZAMECNIK. 1958. J. Biol. Chem. **231:** 241.
3. HÜLSMANN, W. C. & F. LIPMANN. Biochim. et Biophys. Acta. In press.
4. KIRSCH, J. F., P. SIEKEVITZ & G. E. PALADE. 1960. J. Biol. Chem. **235:** 1419.
5. KORNER, A. 1959. Biochim. et Biophys. Acta. **35:** 554.
6. LITTLEFIELD, J. W. & E. B. KELLER. 1957. J. Biol. Chem. **224:** 13.

REACTIONS GOVERNING INCORPORATION OF AMINO ACIDS INTO THE PROTEINS OF THE ISOLATED CELL NUCLEUS

V. G. Allfrey, J. W. Hopkins, J. H. Frenster, A. E. Mirsky
The Rockefeller Institute, New York, N. Y.

The experiments to be described deal with the mechanism of protein synthesis in isolated cell nuclei, particularly in nuclei prepared from the calf thymus gland. The selection of thymus as the tissue of choice for such studies dates back to earlier experiments in which it was found that calf thymus nuclei can withstand an isolation in isotonic sucrose solutions containing a small amount of $CaCl_2$.[1,2] Thymus nuclei so prepared retained their soluble proteins and nucleoproteins, and even low-molecular weight compounds such as nucleotides and free amino acids were retained.[3,4,5]

Tests for nuclear purity. Under the light microscope the thymus nuclear preparations, stained or unstained, appear remarkably clean. Contamination consists of a few whole cells, occasional red cells, and a bit of cytoplasm that here and there adheres to some of the nuclei. The extent of cytoplasmic contamination (as estimated from nucleic acid analyses and tests for the activity of enzymes known to be localized in the cytoplasm) is found to be well below 10 per cent. Recent visual estimates of whole-cell contamination in stained nuclear preparations place the number of intact cells at 3 per cent.[6] This is in accord with earlier whole-cell counts made under the electron microscope.[7] When necessary, most of these few remaining cells can be removed by centrifugation through a Ficoll density gradient.[8]

Nuclear function. It was soon discovered that these isolated nuclear fractions are capable of at least three types of synthetic reactions: (1) the uptake of amino acids into nuclear proteins,[3,9] (2) the synthesis of adenosine triphosphate and other energy-rich phosphate bonds,[4,10] and (3) the incorporation of a variety of purine and pyrimidine precursors into both ribonucleic and deoxyribonucleic acids.[3,11-14]

Many of these reactions have subsequently been found to bear some relationship to each other. Some of the interrelationships that cast some light on the mechanisms of nuclear protein synthesis will now be discussed.

Amino acid incorporation. The over-all process of amino acid incorporation into the mixed proteins of the nucleus can be observed when nuclei are incubated *aerobically* at 37° C. in a buffered sucrose medium that contains radioactive amino acids.[9,3] FIGURE 1 shows the time course of incorporation of lysine-2-C^{14}, alanine-1-C^{14}, and glycine-1-C^{14}. Curves similar to those shown for alanine and lysine are obtained with C^{14}-leucine, phenylalanine, tryptophan, and valine. In the figure the specific activity of the total mixed proteins of the nucleus is plotted against the time of incubation. A "lag" phase of 5 to 15 min. is characteristically observed, followed by a rapid incorporation of label into nuclear protein. This incorporation is essentially irreversible and, if it is all assumed to represent protein synthesis, it can be calculated that 30 mg. of nuclei (10^9 nuclei) are synthesizing about 8 μg. of protein per hour. It should be stressed, however, that not all the nuclei are incorporating amino acid at the same rate. Marked differences in the activity of individual nuclei

are evident in autoradiographs prepared after incorporation has occurred.[6,3] Nevertheless, the extent of the synthesis can be better appreciated from the calculation that, on the average, each nucleus is synthesizing 22 molecules of protein (of average molecular weight 50,000) every second.

The incorporation of amino acids into the proteins of isolated nuclei takes place only under aerobic conditions. The reason for this became clear when other evidence accumulated relating nuclear protein synthesis to a second oxy-

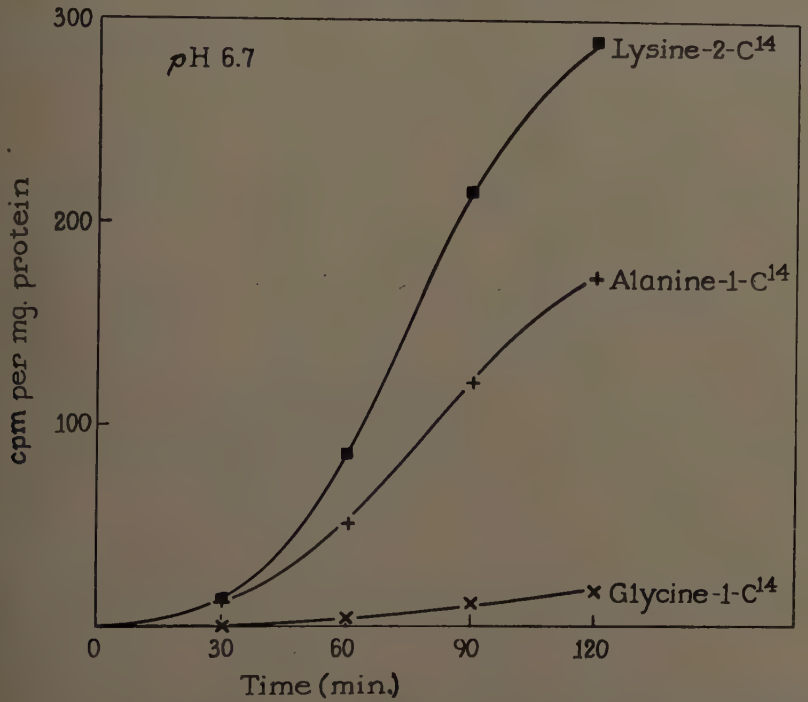


FIGURE 1. The time course of incorporation of C^{14}-labeled amino acids into the proteins of thymus nuclei during incubation *in vitro*. The specific activity of the total mixed protein of the nucleus is plotted against the time of incubation. In this plot the data have been "normalized" to show the relative uptakes for equivalent amounts of the different amino acids when each has the same specific activity in millicuries per millimole.

gen-dependent nuclear function, namely the synthesis of adenosine triphosphate.

Nuclear phosphorylation. Isolated thymus nuclei are capable of phosphorylating the AMP which they contain to form ATP.[4,10] This process, although aerobic, differs in several important respects from the process of oxidative phosphorylation observed in suspensions of isolated mitochondria. For example, it is not affected by such inhibitors of mitochondrial phosphorylation as dicumarol, methylene blue, Ca^{++} ions, or an atmosphere of 95 per cent carbon monoxide-5 per cent oxygen. A detailed comparison of the phosphorylating systems in the nucleus and in mitochondria will not be made here, but it should be stressed that nuclear ATP synthesis is DNA-dependent.[10,15]

The removal of the DNA with pancreatic deoxyribonuclease blocks subsequent ATP synthesis, but restoring the DNA or substituting polyanionic molecules for it allows phosphorylation to proceed. These findings stress the importance

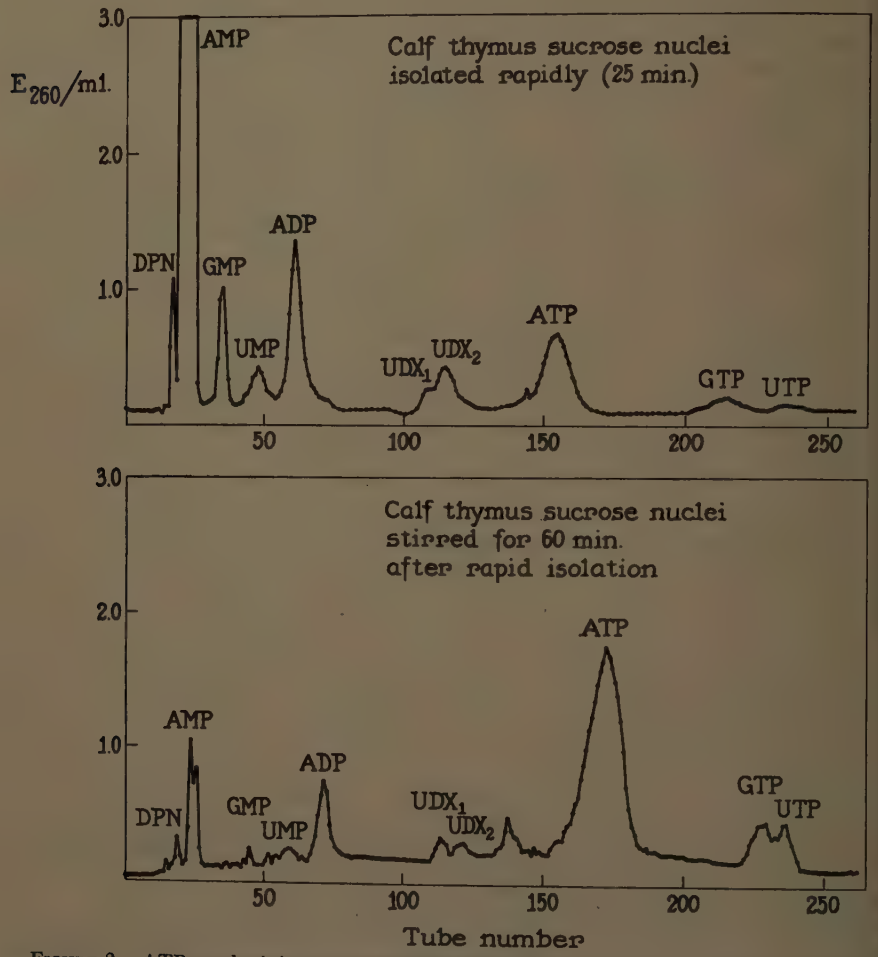


FIGURE 2. ATP synthesis in suspensions of isolated thymus nuclei. The chromatographic separation of ATP and other nucleotides is shown by plotting the optical density of the effluent (E_{260}) against the volume of solution passing through a column of Dowex-1. Note that in the upper curve, AMP is present in large amounts while the ATP peak is small. The lower curve shows that after aeration the ATP content is high.

of the negative electrical charge on DNA in the regulation of nuclear metabolism.[16]

Experimentally, the process of ATP synthesis in the nucleus can be followed by chromatographic isolation of the different nucleotides (FIGURE 2) and by the incorporation of P^{32}-orthophosphate into the terminal phosphate groups of the nucleotides.[8]

The role of ATP in nuclear protein synthesis. A study of the comparative effects of different inhibitory compounds on nuclear protein and ATP synthesis showed that all agents that block the generation of ATP by the nucleus also block amino acid incorporation into its proteins. This suggested that the amino acid uptake is ATP-dependent. Supporting evidence for this assumption came from other experiments in which the nucleotides were selectively removed from the nucleus by treatment with *acetate* buffers at *p*H values below 5.9.[4] In such nuclei there was a clear parallelism between the loss of nucleotides and the decrease in amino acid incorporation. (Using citrate or other buffer systems that do not extract nuclear ATP as "controls," it could be shown that nuclei could recover from exposure to the low *p*H values employed in the acetate system of nucleotide removal.)

Direct evidence for the role of ATP in nuclear protein synthesis is obtained from the study of the amino acid-activating enzymes in nuclear extracts. This aspect of protein synthesis in the nucleus is discussed in detail below.

Stages in Nuclear Protein Synthesis

Amino acid transport and the sodium ion requirement. One of the most striking aspects of amino acid incorporation by isolated cell nuclei is the dependence of protein synthesis on the presence of sodium ions.[3] This requirement has been observed for all of the amino acids tested thus far; for most of them the level of uptake into protein is doubled by adding Na+ (as the chloride) to the incubation medium. The incorporation of alanine-1-C^{14} is especially responsive to the presence of sodium ions. FIGURE 3 shows that the level of uptake in a medium containing sodium (at a concentration of 0.04 *M*) is 10 times that observed with an equivalent amount of potassium. The figure also shows that although lithium is partly effective as a sodium substitute neither potassium, caesium, nor rubidium have any stimulatory effect on C^{14}-alanine incorporation.

The reason why the incorporation of amino acid into nuclear protein increases when sodium ions are added to the incubation medium has now been made clear: amino acid transport into the nucleus appears to be strongly sodium-dependent. This was demonstrated in experiments in which the ionic environment of the nucleus was varied and the penetration of free amino acids into the nuclear amino acid "pool" was measured. A comparison of the rate of alanine-1-C^{14} transport into nuclei in sodium- and potassium-containing media is shown in FIGURE 4. It is clear that the amount of amino acid entering the pool is much greater in a medium containing sodium ions.

That the transport of amino acids into isolated nuclei is not a simple diffusion process has been shown by other experiments testing both the D- and the L-isomers of different amino acids. These experiments will not be described here, but competition studies indicate that the transport process is specific for the L-isomer of the amino acid and that the unnatural D-isomer is not actively taken up.[17]

It should be mentioned in passing that other synthetic processes in the nucleus may or may not show a sodium-ion dependence. A sodium requirement has been found for the uptake of adenine- or adenosine-8-C^{14} into nuclear

RNA and for the incorporation of thymidine-2-C^{14} into DNA. Here, too, the results can be interpreted in terms of a specific stimulation by sodium ions of the transport of both adenosine and thymidine into the nucleus. On the other

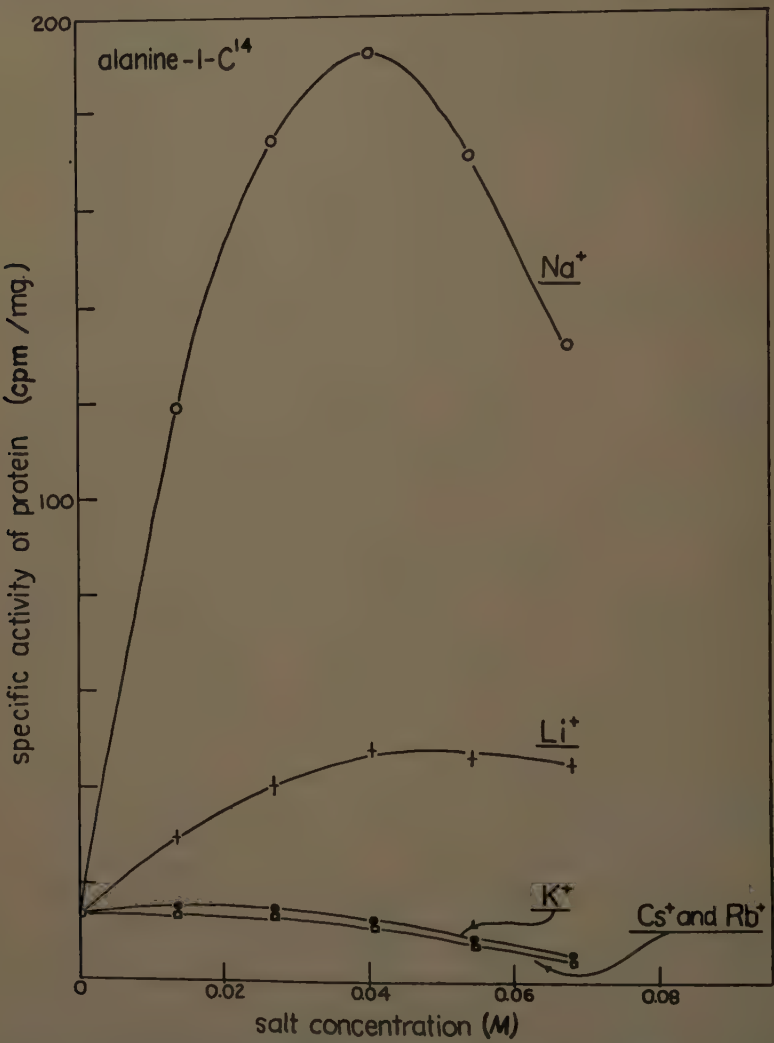


FIGURE 3. The effect of adding different monovalent cations (as chlorides) on the incorporation of C^{14}-alanine into the proteins of isolated thymus nuclei. The specific activity of the nuclear protein after 60 min. incubation is plotted against the salt concentration in the medium.

hand, no specific ionic requirement is observed for orotic acid-6-C^{14} incorporation into the uridylic acid of nuclear RNA. This fits in with observations showing that orotic acid penetration into nuclei is not specifically promoted by sodium ions.[17]

Sodium-ion dependence is very useful as a test indicating that the uptake

observed in a given protein fraction actually occurs in the nucleus and is not due to contaminating cytoplasmic particles. This test owes its validity to the fact that amino acid incorporation in all of the cytoplasmic systems so far studied is potassium-dependent.[18]

The transport experiments have made it clear that one of the first requisites for nuclear protein synthesis, namely the accumulation of free amino acids at the site of synthesis, is dependent upon the presence of sodium ions. Other

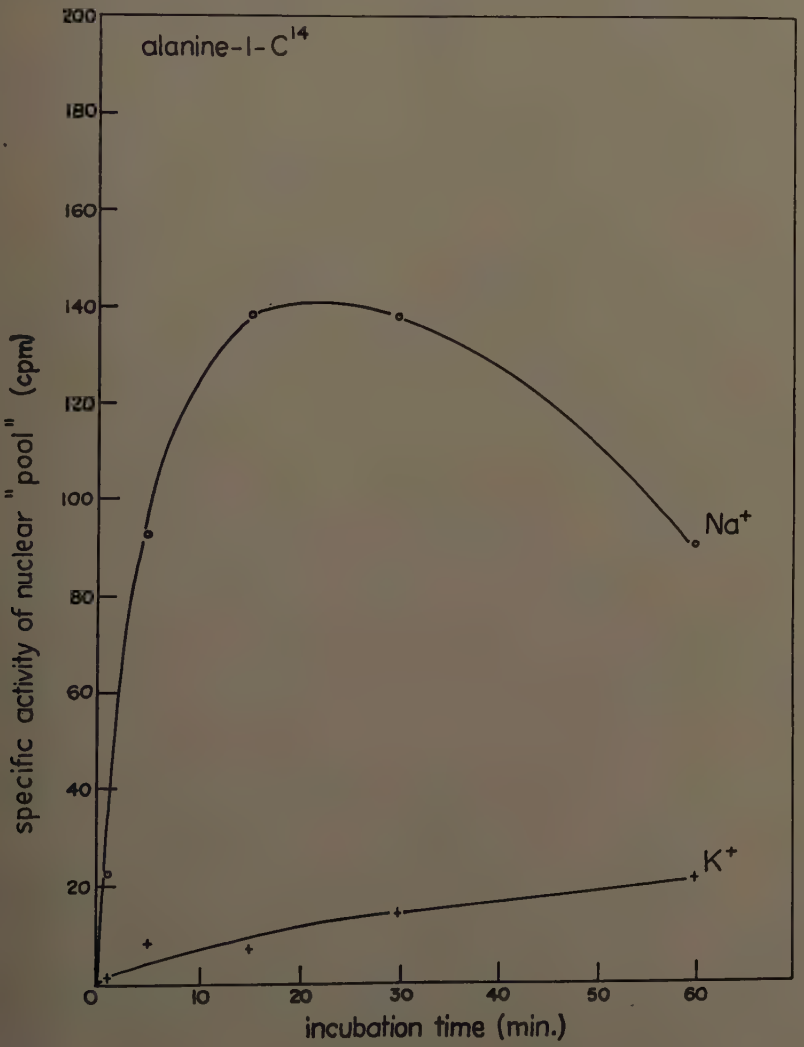


FIGURE 4. The specific effect of Na+ ions in promoting the entrance of amino acids into the isolated nucleus. The specific activity of the free amino acid in the "acid-soluble" nuclear extract is plotted against the time of incubation. Nuclei were incubated at 37° C. with alanine-1-C^{14} for the indicated periods, then centrifuged down in the cold and extracted with cold 2 per cent $HClO_4$.

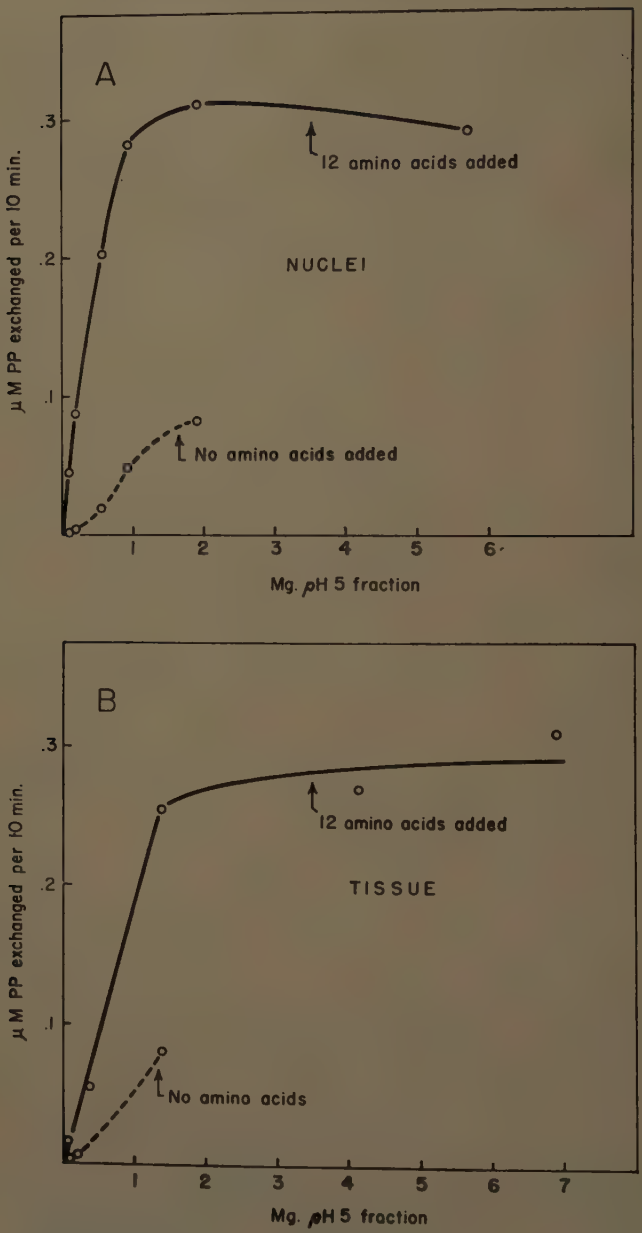


FIGURE 5. The assay of amino acid activating enzymes in *p*H 5 fractions of thymus nuclei and tissue. Increasing amounts of the *p*H 5 enzyme fraction promote the incorporation of P^{32}-pyrophosphate into ATP (ordinate). The addition of a mixture of 12 L-amino acids (upper curves) stimulates the ATP-PP^{32} exchange.

tests have been carried out to see whether later stages in nuclear protein synthesis are also sodium-ion dependent. The results have been negative with respect to amino acid activation and the subsequent step of amino acid transfer to carrier ribonucleic acids (see below).

Amino acid activation. Beginning with the work of Hoagland,[19] evidence has been obtained in many laboratories to indicate that the first step in the sequence of protein-synthetic reactions involves the enzymatic activation of the carboxyl groups of the amino acids according to the reaction:

amino acid + ATP + activating enzyme $\rightleftharpoons$ enzyme-AMP-amino acid + PP

The reaction requires ATP, pyrophosphate is split off, and the amino acyl-adenylate compounds that are formed remain bound to the enzyme.[20]

The assay for enzyme activity is conveniently carried out by studying the reverse reaction, using radioactive pyrophosphate and measuring the incorporation of P^{32} into ATP. This incorporation is dependent upon the presence of free amino acids in the reaction mixture.

Using this test for pyrophosphate-ATP exchange, it can be shown that neutral extracts of thymus nuclei contain amino acid activating enzymes.[21] The methods used in the preparation and testing of the enzymes have been described elsewhere,[21] but it should be mentioned that the active fraction is conveniently prepared by precipitation at *p*H 5.2, as in cytoplasmic systems.

The amino acid dependence of ATP-PP^{32} exchange in *p*H 5 fractions prepared from thymus nuclei isolated in sucrose solutions is shown in FIGURE 5 (together with experimental data obtained from *p*H 5 enzymes prepared from whole, fresh tissue). The amino acid dependence is clear in both cases, and the small amount of exchange observed in the absence of added amino acids can be reduced still further by dialysis of the enzyme preparation to remove residual free amino acids. Alternatively, the residual amino acids can be removed by adsorption on Norite A.[22] The ATP-PP^{32} exchange occurs on the addition of mixtures of amino acids or with individual amino acids tested separately. A survey of the effects of different amino acids is summarized in TABLE 1. Thus far, at least 15 L-amino acids have been found to promote the exchange, and there is no activation of the D-isomers of the amino acids. The rate of exchange depends on the amino acid concentration of the incubation medium; the effect is illustrated for L-leucine in FIGURE 6.

Evidence for nuclear localization of the activating enzymes. A number of tests have been made to verify the finding that amino acid-activating enzymes actually occur in the cell nucleus and to eliminate the possibility that their presence in isolated nuclei is an artefact resulting from the adsorption of cytoplasmic enzymes during the process of isolation.

The best evidence for nuclear localization is obtained by a study of nuclei isolated in nonaqueous media.[23,2] In this method the tissue is rapidly frozen and lyophilized, thus preventing the loss or exchange of water-soluble cell components. After grinding, the nuclei can be isolated by density gradient centrifugation employing mixtures of organic solvents (for example, cyclohexane and CCl_4). Nuclear preparations of high purity have been obtained from a wide variety of animal tissues.[2] The best of these, judging by chemical,

enzymatic, and immunological tests for cytoplasmic contamination, are the nuclei prepared from calf-thymus tissue and from chicken-kidney tissue. These have now been examined for their content of amino acid activating enzymes,[21] and it was found that the nonaqueous thymus nuclei contain slightly more activating activity than the equivalent weight of whole tissue; in kidney nuclei the total activity (in units per microgram of dry weight) is about one

TABLE 1

EFFECT OF AMINO ACIDS ON ATP-PYROPHOSPHATE EXCHANGE CALF THYMUS NUCLEAR *p*H 5 FRACTION

Amino acids promoting exchange	Amino acids without effect in this system
L-Alanine	L-Arginine*
L-Aspartic acid	Glycine*
L-Cysteine	L-Phenylalanine*
L-Glutamic acid	All D-Amino acids
L-Histidine	
L-Isoleucine	
L-Leucine	
L-Lysine	
L-Methionine	
L-Proline	
L-Serine	
L-Threonine	
L-Tryptophan	
L-Tyrosine	
L-Valine	

* Found in low concentration in 0.05 M KCl supernatant before precipitation of *p*H 5 fraction.

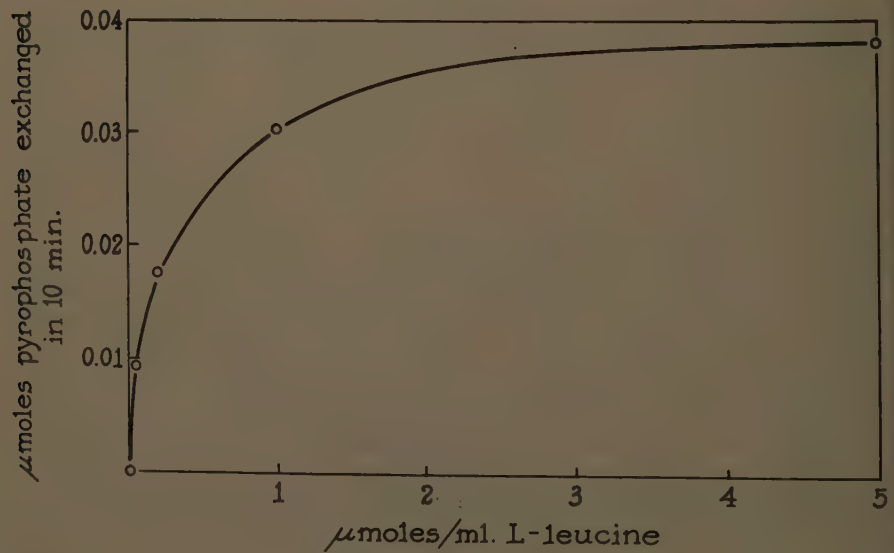


FIGURE 6. Assay of the leucine-activating enzyme in the nuclear *p*H 5 fraction. The rate of ATP-pyrophosphate exchange is plotted against the concentration of L-leucine in the medium.

half that of the tissue. In both cases, it is clear that this amount of activating enzyme cannot be explained in terms of cytoplasmic contamination of the isolated nuclei.

We have recently observed the amino acid activation reaction in *p*H 5 fractions prepared from calf-liver nuclei following their isolation in dense sucrose solutions by the method of Chauveau *et al.*[24] These nuclei, too, are characterized by a high degree of purity. In a recent communication Webster[25] has reported the presence and partial purification of an alanine-activating enzyme in nuclei prepared from pig liver.

Thus the presence of amino acid-activating enzymes promises to be a general characteristic of nuclei that are actively synthesizing protein. It remains to be seen whether such enzymes occur in nuclei that do not carry out an appreciable protein synthesis, such as avian erythrocyte or Arbacia sperm nuclei.

Amino acid transfer to nuclear ribonucleic acids. In many cytoplasmic systems it has been shown[26-30] that the sequel to the activation of the amino acids is their transfer to "soluble" ribonucleic acids according to the reaction:

$$\text{enzyme-aminoacyl AMP} + \text{sRNA} \rightleftharpoons \text{aminoacyl-sRNA} + \text{enzyme} + \text{AMP}$$

The evidence will now be presented for a similar reaction in isolated thymus nuclei and in nuclei prepared from calf-liver tissue.

In intact thymus nuclei the transfer of amino acid to ribonucleic acid is readily demonstrated by incubating the nuclei in the presence of radioactive amino acid and adding chloramphenicol to inhibit later stages in protein synthesis. (The inhibitory effect of chloramphenicol on protein synthesis in thymus nuclei was first pointed out by Breitman and Webster.[31] We have verified their findings and further established that this antibiotic does not affect either nuclear ATP synthesis or the amino acid activation reaction.[21]) Following incubation, ribonucleic acid can be isolated by the phenol procedure[32] and its radioactivity measured.

The results of such an experiment are shown in FIGURE 7. Here the specific activity of nuclear protein and nuclear RNA are plotted against the time of incubation in the presence of leucine-1-C^{14}. It is clear that the addition of chloramphenicol effectively stops amino acid uptake into the protein of the nucleus. This permits the isolation of RNA with no danger of contamination by radioactive protein. Evidently chloramphenicol does not prevent the transfer of amino acid to nuclear ribonucleic acid, thus permitting the isolation of the radioactive complex.

Similar results have recently been obtained using another antibiotic: puromycin. This approach was suggested by the work of Yarmolinsky and de la Haba,[33] who have shown that puromycin in very low concentrations (200 mμM) can completely inhibit amino acid incorporation into the proteins of rat-liver extracts. Like chloramphenicol in the nuclear system, puromycin does not affect ATP synthesis, amino acid activation, or amino acid transfer to soluble RNA in the liver extracts.[33] Its structure is remarkably similar to that proposed for the amino acyl-RNA complex (FIGURE 8). The inhibition of protein synthesis in nuclei requires considerably more puromycin than is effective in liver cytoplasm; we have found that 20 μM reduces the nuclear uptake of alanine-1-C^{14}, glycine-1-C^{14}, and leucine-1-C^{14} by about 50 per cent. At a puromycin concentration of 80 μM, the incorporation of these amino acids is

only 15 per cent of that observed in the absence of the antibiotic. It is of interest that, even at this relatively high level, puromycin has no appreciable effect on over-all nuclear RNA synthesis; the incorporation of adenosine-8-C^{14} into RNA is unimpaired, and guanine-8-C^{14} and orotic acid-6-C^{14} uptakes are reduced only slightly.

Considering its structure, the effectiveness of puromycin as an inhibitor of nuclear protein synthesis can be viewed as the result of a close competition be-

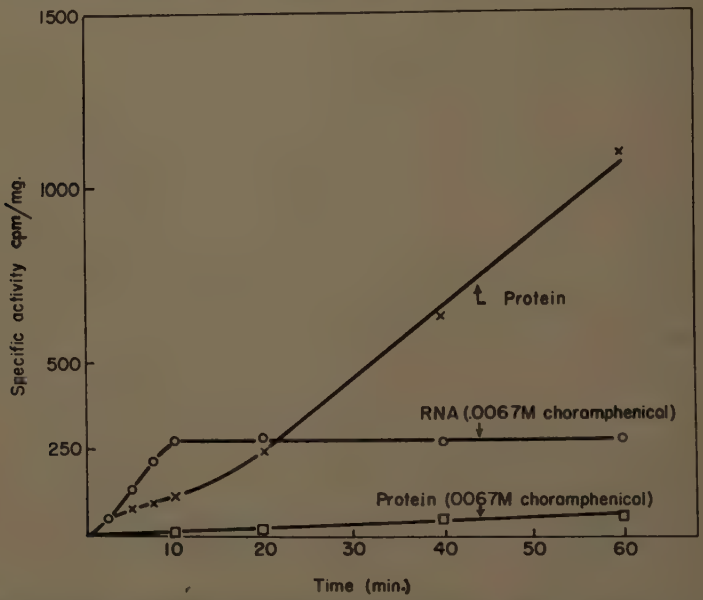
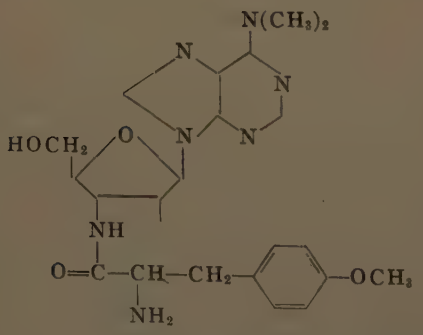


FIGURE 7. The uptake of leucine-1-C^{14} into the protein and "transfer" RNA of isolated nuclei. The specific activity of the protein (cpm/mg.) is plotted in the upper curve. The incorporation into protein is markedly inhibited in the presence of chloramphenicol (lower curve). The uptake of amino acid into the RNA is the same in chloramphenicol-treated nuclei as in "controls."



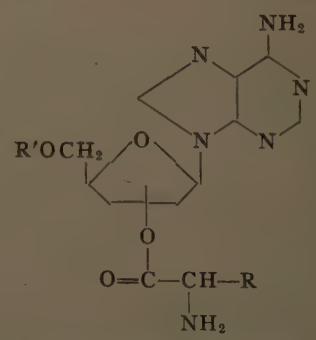


FIGURE 8. A comparison of the structures of puromycin and amino acyl-transfer RNA (from Yarmolinsky and de la Haba).[33]

tween the antibiotic and the natural amino acyl-RNA complex. The fact that nearly all of the C^{14}-amino acid uptake into protein is blocked by adding puromycin is strong presumptive evidence for the view that *most* of the incorporation of amino acid into protein proceeds through an RNA-amino acyl intermediate. The structure of the amino acyl-RNA complex in the nucleus is discussed in some detail below.

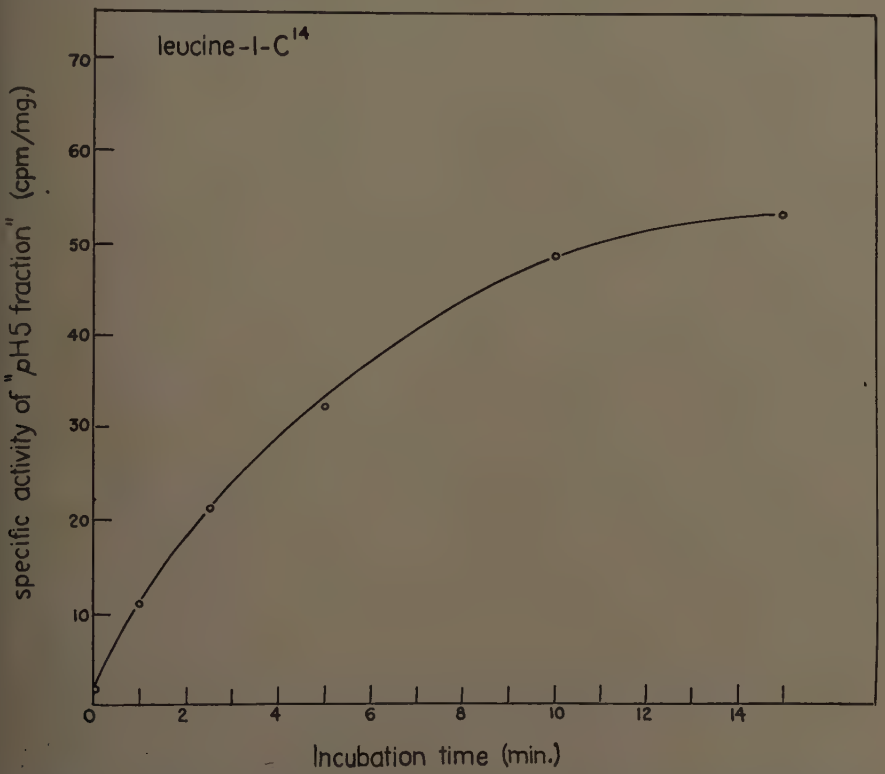


FIGURE 9. The time course of C^{14}-leucine incorporation into the transfer RNA of the nuclear *p*H 5 fraction. The *p*H 5 fraction was incubated for the indicated periods in the presence of ATP and leucine-1-C^{14}.

Amino acyl-RNA formation in the isolated pH 5 fraction. The labeling of nuclear RNAs with radioactive amino acids can be carried out either in the intact nucleus or in an isolated system containing nuclear activating enzymes, RNA, ATP, and the C^{14}-amino acid in suitable concentrations. The activating enzymes prepared at *p*H 5 by the usual procedure contain a few percent of nuclear RNA that can itself accept the isotopic amino acid, but a higher level of amino acid transfer to ribonucleic acid is obtained when more nuclear RNA (prepared by the phenol procedure) is added to the system.

The time course of the reaction is illustrated in FIGURE 9, which shows the extent of leucine-1-C^{14} incorporation into the isolated *p*H 5 fraction. That the uptake actually represents incorporation of amino acid into RNA can be

shown in several ways; (1) the RNA can be subsequently prepared by the phenol procedure and shown to contain the C^{14}-amino acid, (2) all of the radioactivity can be removed from the labeled *p*H 5 fraction by treatment with ribonuclease, and (3) the presence of 10 μg. of ribonuclease during the incubation completely prevents the incorporation of amino acid. It has also been found that all of the counts are removed by alkaline digestion of the RNA or by treatment with hot trichloroacetic acid to remove nucleic acid. Tests of

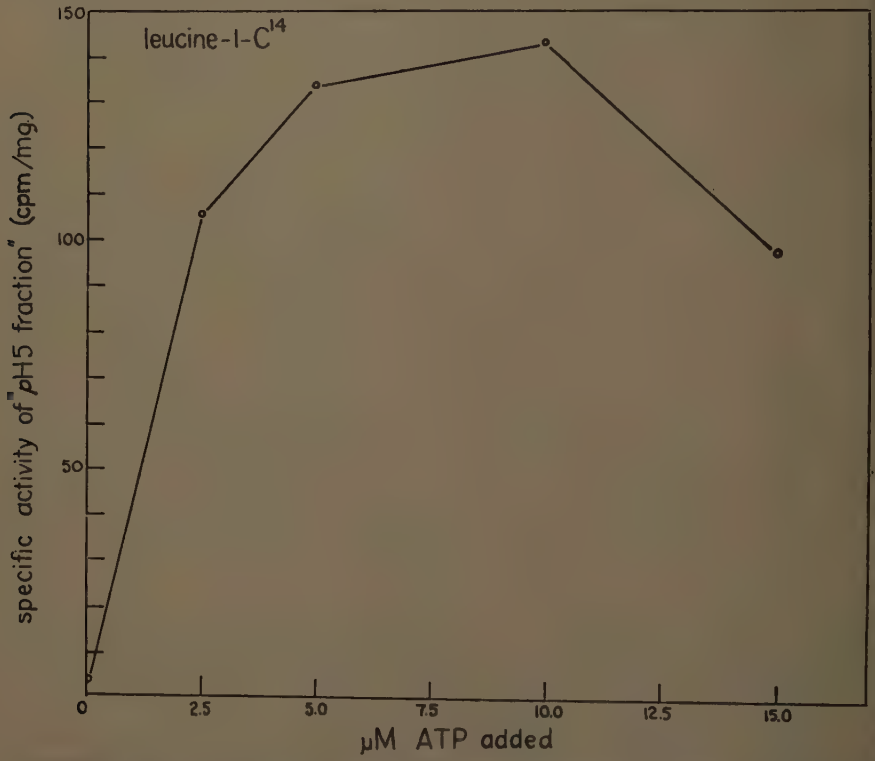


FIGURE 10. The relationship between ATP concentration and C^{14}-leucine uptake into the transfer RNA of the nuclear *p*H 5 fraction.

the protein residue after these treatments confirm that there is no significant labeling of the protein in the *p*H 5 fraction.

The formation of the amino acyl-RNA complex requires the presence of ATP. FIGURE 10 shows that this substance is most effective at concentrations between 5 to 10 μM per ml. The addition of cytidine triphosphate (CTP) to the incubation mixture does not promote a higher rate of incorporation of leucine-1-C^{14} into nuclear RNA if the *p*H 5 fraction is freshly prepared. However, if the fraction is pre-incubated at 37° C. before testing its capacity for amino acid incorporation, then the addition of CTP stimulates the reaction.

Tests for specificity. It is a matter of some interest to know whether the "transfer" RNA of the thymus nucleus can accept amino acids from activating

enzymes prepared from the nuclei of other tissues. A few preliminary experiments have been carried out to test for such cross-specificity. The results shown in TABLE 2 show that guinea pig-liver enzymes can transfer C^{14} leucine to the RNA's prepared from either calf-thymus nuclei or calf-liver nuclei. Conversely, the activating enzymes of the thymus nucleus will transfer leucine-1-C^{14} to both thymus nuclear RNA and calf-liver nuclear RNA. Thus there is no evidence of nuclear enzyme or RNA specificity as far as the tests go. However, it should be mentioned that Webster[25] has recently reported that the alanine-activating enzymes prepared from pig-liver nuclei and cytoplasm do show differences in their capacities to transfer alanine to different RNA preparations. It follows that the lack of specificity observed in C^{14}-leucine labeling of different nuclear RNA's may not be an adequate test of this interesting problem. Other experiments using C^{14}-valine, lysine, and methionine are now in progress.

Nucleolar RNA. Not all of the ribonucleic acid fractions of the thymus nucleus function as amino acid acceptors in the transfer reaction. The RNA

TABLE 2

CROSS-SPECIFICITY IN LEUCINE-1-C^{14} TRANSFER TO DIFFERENT NUCLEAR RNAS

Source of RNA	Source of pH 5 enzymes	Total counts transferred to RNA
Calf Thymus nuclei	Calf-Thymus nuclei	647
Calf Liver nuclei	Calf-Thymus nuclei	692
Calf Thymus nuclei	Guinea-pig liver*	477
Calf Liver nuclei	Guinea-pig liver*	406

* We are indebted to Jack F. Kirsch of The Rockefeller Institute for the gift of the activating enzymes of guinea pig liver.

preparations tested in the above experiments were obtained by the phenol method and certainly represent a diverse mixture of different molecular species. However some selection has occurred, for we have found that the phenol procedure, as usually carried out, does not permit the isolation of nucleolar RNA.

For this reason the capacity of nucleolar RNA to function as an amino acid acceptor had to be tested separately. In this experiment the incorporation of leucine-1-C^{14} was allowed to proceed in intact nuclei and the RNA was subsequently fractionated on the basis of its solubility in salt solutions by methods described elsewhere.[11] The results showed that the insoluble "nucleolar" RNA fraction does not incorporate significant amounts of labeled amino acid; most of the leucine-C^{14} is found in the readily extractable RNA of the nucleus.[21] This evidence for the low activity of nucleolar RNA with respect to amino acid incorporation is in accord with autoradiographic evidence that shows that amino acid uptake into the nucleolus is much lower than that observed on other regions of the nucleus.[34]

Is amino acid transferred to DNA in the nucleus? The incorporation of amino acids into the proteins of thymus nuclei is blocked when the nuclei are first treated with deoxyribonuclease.[3] Amino acid uptake can be restored by restoring thymus DNA, but RNA, other polynucleotides, and even synthetic

polyanions (such as polyethylene sulfonate) can also restore C^{14} uptake.[15,16] These restoration experiments with synthetic polyanions made it very unlikely that DNA is directly involved in protein synthesis. Nevertheless, a direct test of the possibility that DNA might function as an intermediate acceptor of amino acids has been carried out, with negative results. Nuclei were incubated with leucine-2-C^{14} and the DNA subsequently isolated by the sodium dodecyl sulfate method B of Marko and Butler.[35] No significant C^{14} incorporation into the DNA was observed.

Properties of the amino acyl-RNA complex. It has already been pointed out that the RNA associated with the nuclear *p*H 5 fraction can be labeled by an *in vitro* incubation with ATP and leucine-1-C^{14}. The chemical nature of the amino acyl-RNA complex has been studied using this reaction as a preparative procedure. The product leucyl-RNA has a number of properties that indicate a close resemblance to the amino acyl-RNA complexes studied in cytoplasmic systems:[26-30] (1) it is largely nondialyzable, (2) it is stable in dilute acid, (3) upon hydrolysis on 0.005 *N* NaOH and subsequent chromatography on paper[36] the C^{14} activity is removed from the RNA and recovered as free leucine, and (4) a short digestion with ribonuclease (10 μg, 15 min. at 37° C.) separates the radioactivity from the acid-precipitable fragments.

It can be shown that the enzymatic breakdown of the nuclear amino acyl-RNA complex releases the amino acid bound to a nucleoside. Following the procedure of Zachau *et al.*,[37] we have separated this amino acyl-nucleoside by ionophoresis on filter paper. Subsequent treatment of the spot with dilute alkali releases the activity as free C^{14}-leucine. Alternatively, the amino acyl-nucleoside complex can be isolated by chromatography on paper.[38] The C^{14} activity and the UV absorption migrate together. The separation achieved is shown in FIGURE 11, in which the radioactivity is plotted against the distance of migration. For purposes of comparison, the markers leucine-1-C^{14} and adenosine-8-C^{14} are run on the same sheet. It is clear that the R_f of the amino acyl-nucleoside allows a clear separation from these markers. Again, alkaline treatment of the spot releases all of the C^{14} activity as free leucine.

Some preliminary attempts have been made to identify the base in the nucleoside prepared in this way from nuclear RNA. The problem here is to get the material in sufficient quantity for clear-cut spectral analysis; to date, a positive identification has not been made. On the basis of results obtained in cytoplasmic systems,[37] we expect the base to be adenine. This would then be an indication that the "transfer" RNA of the nucleus probably has a terminal nucleotide sequence (adenylyl-cytidylyl-cytidylyl-) as in cytoplasmic systems.[39] However, some of the nuclear ribonucleic acid appears to contain a guanylic-uridylic sequence at the end of the chain (the evidence for this will be presented elsewhere[40]). Both of these nucleotides are engaged in a "turnover" reaction, and it is tempting to speculate that this sequence also is important in some phase of protein biosynthesis.

Amino acid incorporation into ribonucleoprotein particles from the cell nucleus. In earlier experiments it was observed that the extraction of isolated thymus nuclei with neutral buffer solutions yielded fractions containing ribonucleic acid and protein. Amino acid incorporation studies showed that the protein

in these fractions was radioactive and that it had a higher specific activity after labeling experiments than the average mixed protein of the nucleus.[3]

More recent studies of such nuclear extracts have revealed that they contain a heterogeneous population of ribonucleoprotein particles. By differential ultracentrifugation these particles have been fractionated into classes that differ in their chemical composition and metabolic acitvity. The details of this centrifugal fractionation have been published elsewhere.[41]

Electron microscopy of the separate fractions reveals the presence of large numbers of dense granules of 100 Å diameter. Particles of this size are clearly

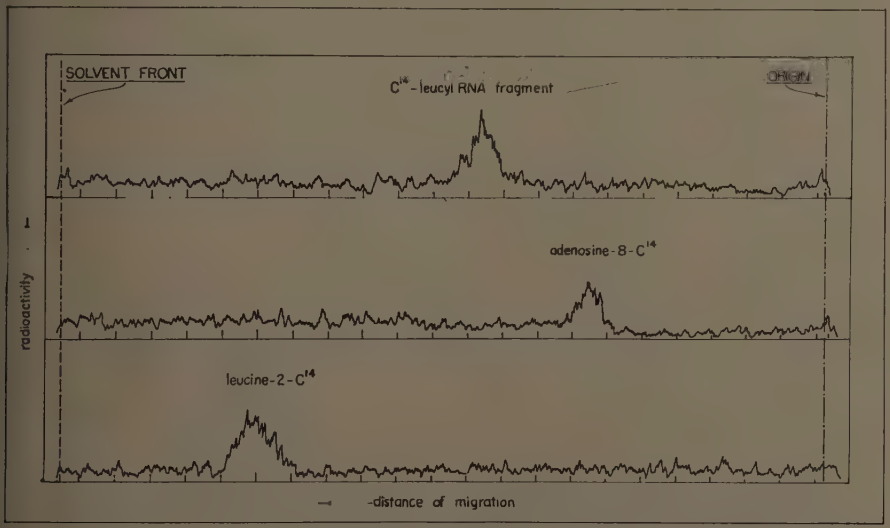


FIGURE 11. Separation of the amino-acyl-nucleoside fragment from nuclear transfer RNA. The RNA was labeled with leucine-1-C^{14} by an *in vitro* incubation with the nuclear *p*H 5 enzymes. The terminal nucleoside containing bound C^{14}-leucine was released by a brief digestion with ribonuclease and separated by chromatography on paper. The location of the fragment was detected by scanning for C^{14}-activity. The radioactivity is plotted against the distance of migration along the paper. The position of C^{14}-adenosine and C^{14}-leucine "markers" is indicated in the lower curves.

discernible in sections of intact thymus nuclei and have also been observed in nucleoli,[42-44] in chromosomal loops,[45] in the Balbiani rings of certain chromosomes,[46] and throughout the nuclear sap[47] in other tissues.

Tracer experiments have shown that the nuclear ribonucleoprotein particles, like those of the cytoplasm, are actively engaged in protein synthesis. Their synthetic activity within the nucleus can be followed by incubating isolated nuclei in the presence of C^{14}-labeled amino acids and subsequently isolating the RNP fractions.[41] These uptake experiments indicate that the different particle fractions, separated on the basis of different sedimentation characteristics, also differ in their metabolic activity.

It should be stressed that two types of evidence testify to the nuclear localization of this activity:

(1) The incorporation of amino acids into the protein of the particles requires

the presence of sodium ions. As mentioned above, this ionic requirement for sodium, rather than potassium, indicates nuclear localization.

(2) Pretreatment of the nuclei with deoxyribonuclease prevents any subsequent uptake of C^{14}-amino acids into the particles. This DNA dependence shows that the incorporation of amino acid into the RNA particles requires the *nuclear* phosphorylating system as a source of ATP.[10]

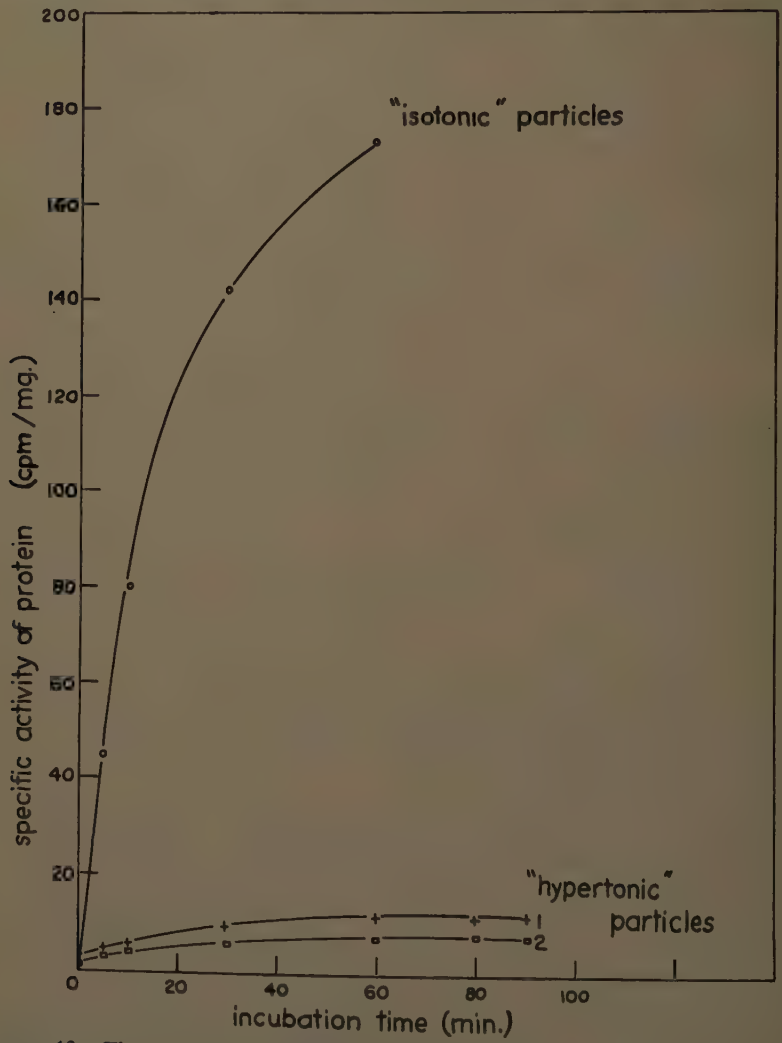


FIGURE 12. The time course of incorporation of leucine-1-C^{14} into isolated ribonucleoprotein particles prepared from thymus nuclei. The incubation medium contained (in addition to the particles) nuclear *p*H 5 enzymes, ATP plus an ATP-generating system, and GTP. The upper curve shows the specific activity of the particle protein when the RNP particles were isolated under isotonic conditions. The lower curves show that particles prepared in hypertonic sucrose solutions are not capable of amino acid uptake under these conditions. In curve 1, the *p*H 5 fraction was prepared using isotonic media; in curve 2, the *p*H 5 fraction was prepared in hypertonic sucrose solutions.

The experiments described above have been concerned with amino acid uptake in RNP particles within the cell nucleus. It has recently been found that the isolated ribonucleoprotein particles can themselves incorporate amino acids in a nucleus-free system provided they are properly supplemented. This reaction requires, in addition to the particles, ATP (and an ATP-generating system), the nuclear *p*H 5 enzymes, and guanosine triphosphate (GTP). The GTP requirement for amino acid incorporation into RNP particles has also been observed in cytoplasmic systems.[48]

The time course of leucine-1-C^{14} incorporation into the proteins of isolated nuclear RNP particles is shown in FIGURE 12. It is of interest that the reaction takes place only if the particles have been prepared under isotonic conditions (upper curve); ribonucleoprotein fractions isolated in heavy sucrose solutions are completely inert (lower curves). This accords with earlier experiments with intact nuclei, in which it was found that amino acid uptake proceeds best in isotonic sucrose solutions and that protein synthesis is seriously impaired when nuclei are exposed to either hypotonic or hypertonic conditions.

Conclusion

Thus with respect to the pathways of amino acid incorporation into protein, the cell nucleus has much in common with the cytoplasm. Amino acids are activated in the same way, are transferred to "carrier" ribonucleic acids, and make their way into ribonucleoprotein particles. In the nucleus all of these reactions are DNA-dependent because the necessary ATP production is DNA-dependent, but each of the separate reactions involved in the early stages of protein synthesis can be studied outside of the nucleus when exogenous ATP is supplied.

The results thus far obtained supply only merest suggestions of the over-all synthetic process. It remains to be determined how the sequence of amino acids in newly synthesized protein is directed in the ribonucleoprotein particle and how the structure of this complex RNP unit is itself determined by the DNA in the chromosomes.

References

1. STERN, H. & A. E. MIRSKY. 1953. J. Gen. Physiol. **37:** 177.
2. ALLFREY, V. G., H. STERN, A. E. MIRSKY & H. SAETREN. 1952. J. Gen. Physiol. **35:** 529.
3. ALLFREY, V. G., A. E. MIRSKY & S. OSAWA. 1957. J. Gen. Physiol. **40:** 451.
4. OSAWA, S., V. G. ALLFREY & A. E. MIRSKY. 1957. J. Gen. Physiol. **40:** 491.
5. KAY, E. R. M. 1956. Federation Proc. **15:** 107.
6. FICQ, A. & M. ERRERA. 1958. Exptl. Cell Research. **14:** 182.
7. ALLFREY, V. G. & A. E. MIRSKY. 1955. Science. **121:** 879.
8. ALLFREY, V. G. 1959. Unpublished observations.
9. ALLFREY, V. G. 1954. Proc. Natl. Acad. Sci. U. S. **40:** 881.
10. ALLFREY, V. G. & A. E. MIRSKY. 1957. Ibid. **43:** 589.
11. ALLFREY, V. G. & A. E. MIRSKY. 1957. Ibid. **43:** 821.
12. BREITMAN, T. & G C. WEBSTER. 1959. Nature. **184:** 637.
13. ALLFREY, V. G. & A. E. MIRSKY. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 1325.
14. FRIEDKIN, M. & H. WOOD. 1956. J. Biol. Chem. **220:** 639.
15. ALLFREY, V. G. & A. E. MIRSKY. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 981.
16. ALLFREY, V. G. & A. E. MIRSKY. 1958. Trans. N. Y. Acad. Sci. Ser. II. **21:** 3.
17. ALLFREY, V. G. & A. E. MIRSKY. 1960. In press.
18. SACHS, H. 1958. J. Biol. Chem. **233:** 643.
19. HOAGLAND, M. B. 1955. Biochim. et Biophys. Acta. **16:** 288.

20. Hoagland, M. B., E. B. Keller & P. C. Zamecnik. 1956. J. Biol. Chem. **218:** 345.
21. Hopkins, J. W. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 1461.
22. Clark, J. M. 1958. J. Biol. Chem. **233:** 421.
23. Behrens, M. 1932. Z. physiol. Chem. **209:** 59.
24. Chauveau, J., Y. Moule & C. Rouiller. 1956. Exptl. Cell Research. **11:** 317.
25. Webster, G. C. 1960. Biochem. Biophys. Res. Comm. **2:** 56.
26. Hoagland, M. B., P. C. Zamecnik & M. L. Stephenson. 1957. Biochim. et Biophys. Acta. **24:** 215.
27. Ogata, K. & H. Nohara. 1957. Biochim. et Biophys. Acta. **25:** 660.
28. Holley, R. W. 1957. Am. Chem. Soc. **79:** 658.
29. Berg, P. & E. J. Ofengand. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 78.
30. Weiss, S. B., G. Acs & F. Lipmann. 1958. Ibid. **44:** 189.
31. Breitman, T. & G. C. Webster. 1958. Biochem. et Biophys. Acta. **27:** 409.
32. Kirby, K. S. 1956. Biochem. J. **64:** 405.
33. Yarmolinsky, M. B. & G. L. de la Haba. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 1721.
34. Carneiro, J. & C. P. Leblond. 1959. Science. **129:** 391.
35. Marko, A. M. & G. C. Butler. 1951. J. Biol. Chem. **190:** 65.
36. Levy, A. L. 1954. Nature. **174:** 126.
37. Zachau, H. G., G. Acs & F. Lipmann. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 885.
38. Roberts, R. B., P. H. Abelson, E. B. Bolton, D. B. Cowie & R. J. Britten. 1955. Carnegie Inst. Wash. Bull. **607.**
39. Hecht, L. I., M. L. Stephenson & P. C. Zamecnik. 1958. Biochem. et Biophys. Acta. **29:** 460.
40. Allfrey, V. G. & A. E. Mirsky. In press.
41. Frenster, J. H., V. G. Allfrey & A. E. Mirsky. 1960. Proc. Natl. Acad. Sci. U. S. **46:** 432.
42. Porter, K. R. 1954. J. Histochem. and Cytochem. **2:** 346.
43. Bernhard, W., A. Bauer, A. Gropp, F. Hagenau & C. Oberling. 1955. Exptl. Cell Research. **9:** 88.
44. Lafontaine, J. G. 1958. J. Biophys. and Biochem. Cytol. **4:** 229.
45. Gall, J. G. 1956. Ibid. **2**(supp.)**:** 393.
46. Beerman, W. & G. F. Bahr. 1954. Exptl. Cell Research. **6:** 195.
47. Callan, H. G. 1956. In Symposium On the Fine Structure of Cells (Leiden). I.U.B.S. Publ. **B21:** 89.
48. Littlefield, J. W. & P. C. Zamecnik. 1957. J. Biol. Chem. **224:** 13.

INTRACELLULAR SITES FOR AMINO ACID INCORPORATION INTO PROTEINS*

Renzo Rendi and Robert C. Warner
Department of Biochemistry, New York University College of Medicine, New York, N. Y.

Early work on the incorporation of labeled amino acids into the proteins of subcellular fractions demonstrated that microsomes play an important role in protein synthesis.[1] Later work has shown that ribosomes are the actual site for the incorporation of amino acids into the ribosomal proteins.[2] More recently, it was observed that other intracellular fractions, for example, nuclei[3,4] and mitochondria,[5,6] are able to incorporate amino acids into their proteins. We therefore wished to determine whether a common particulate component, the ribosome, is involved in the synthesis of proteins in the different fractions.

The ribosomes can be prepared from liver microsomes by treatment with deoxycholate.[2] If this treatment is carried out in the presence of Mg^{++} and lubrol, the ribosomes that are obtained are still able to incorporate labeled amino acids into their proteins *in vitro*.[7] By the same procedure it is possible to prepare ribosomes from isolated mitochondria[8] or nuclei[9] of rat liver. The incorporation into these ribosomes, as into microsomes, is RNase-sensitive. When intact fractions (nuclei and mitochondria) are incubated with a labeled amino acid, the proteins of the ribosomes are more radioactive than those of the other fractions thus far isolated except as discussed below. This labeling is not prevented by RNase, suggesting that the ribosomes are not microsomal contaminations.

When nuclei are incubated *in vitro* with labeled amino acids, the proteins of one of the fractions isolated become more radioactive than those of the nuclear ribosomes. In the case of calf-thymus nuclei the most active fraction is the one bound to the DNA,[3] while in the case of rat-liver nuclei the most active fraction is the residue after *M* NaCl extraction.[9] Because of the difference between the results with rat-liver and calf-thymus nuclei and since in the last case it has been reported that varying the *p*H of the incubation medium changes the pattern of the labeling of the different fractions,[3] it was desirable to know which of these fractions in rat-liver nuclei becomes most radioactive after *in vivo* administration of a labeled amino acid. A short time after an I.P. injection of C^{14}-labeled leucine to rats, the different fractions of the liver nuclei exhibited the pattern of incorporation shown in TABLE 1. It may be seen that the most highly labeled protein was found in the ribosomal fraction. On the other hand, both the residual protein bound to the DNA and the *M* NaCl-insoluble fractions showed a much lower radioactivity. This would suggest that the previously reported data may be due to unphysiological conditions during the *in vitro* incubations.

Evidence for a similarity in structure of the ribosomes present in the different subcellular fractions of rat liver has been obtained by studying the sedimentation properties of the RNA from various fractions. When liver microsomes

* The work reported in this paper was supported in part by Grant NSF-G-9621 from the National Science Foundation, Washington, D.C.

are treated with phenol, the RNA obtained shows two major peaks having sedimentation coefficients of 16S and 23S, respectively (FIGURE 1*a*). The sedimentation coefficients and the relative proportions of these high molecular weight RNA's change with the concentration of Mg^{++}. While the high molecular weight RNA is characteristic of the ribosomes, the cytoplasmic supernatant contains, as is known, an RNA of small molecular size[10] (S-RNA, FIGURE

TABLE 1

IN VIVO LABELING OF THE PROTEINS OF NUCLEAR FRACTIONS OF RAT LIVER AFTER INJECTION OF C^{14}-LEUCINE

	Cpm/mg. protein
H_2O soluble	
high speed supernatant	63
ribosomes	113
M NaCl soluble	
residual protein bound to DNA	52
M NaCl insoluble	45

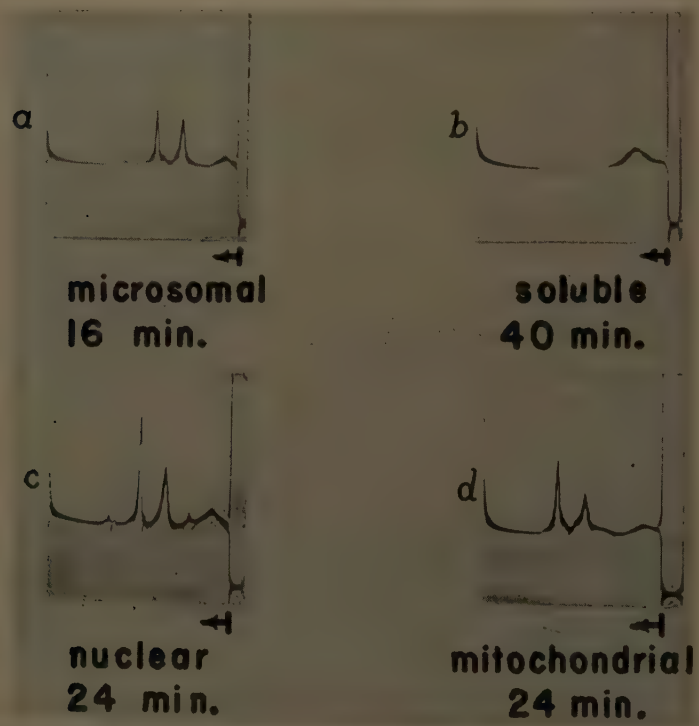


FIGURE 1. Sedimentation patterns of RNA prepared from various cell fractions by the phenol method. The patterns were obtained in 0.02 *M* Tris buffer at *p*H 7.8, containing 0.001 *M* $MgCl_2$ after centrifugation for the indicated time at 59,780 rpm at 20° C. The sedimentation coefficients quoted in the text refer to sedimentation at a concentration of 3 to 4 mg. RNA per ml. The patterns shown are those of RNA from microsomal, nuclear, and mitochondrial ribosomes and of RNA from the supernatant fraction.

1*b*). RNA of rat-liver nuclei, treated with deoxycholate to remove cytoplasmic contaminants, shows the presence of both high molecular RNA's (FIGURE 1*c*). Similar results are obtained for the mitochondrial RNA (FIGURE 1*d*). These observations strongly suggest that the nuclear and mitochondrial ribosomes are similar in structure to the microsomal ribosomes. Since such high molecular RNA's could be isolated from microsomes of various tissues of the rat and from particles from *Azotobacter*, it would seem that the ribosomes of all different cells have a common structure. Similar RNA patterns have been reported for beef-liver microsomes[11] and plant tissues.[12]

Mitochondrial and nuclear RNA not only contain high molecular weight components but, as shown in FIGURE 1*c* and *d*, an RNA with a sedimentation coefficient of 4S, similar to S-RNA. This low molecular weight RNA may be the same as that recently isolated by Hopkins[13] from calf-thymus nuclei that is able to incorporate labeled amino acids. As may be seen in FIGURE 1*a* microsomal ribosomes also contain a small molecular weight RNA that has a sedimentation coefficient of 4S. That this RNA is part of the ribosomes and

TABLE 2

IN VIVO AND *IN VITRO* LABELING OF *p*H 5-RNA AND RIBOSOMAL RNA (4S) WITH C^{14}-LEUCINE

	In vivo	*In vitro*
	Cpm/mg. RNA	
Ribosomal RNA (4S)	4	2200
*p*H 5-RNA	72	2970

is not contaminating cytoplasmic RNA is suggested by the fact that washings and deoxycholate treatments of the microsomes and various other treatments of the ribosomes do not affect substantially its amount relative to the high molecular weight RNA. This RNA has different characteristics from the S-RNA. As shown in TABLE 2, it is able to incorporate labeled amino acids when added to the *p*H 5 fraction of rat liver but, when C^{14}-leucine is injected *in vivo*, no labeling is found in this fraction. It is suggested that this particular RNA represents the S-RNA after the transfer to ribosomes has occurred. Further work is needed to elucidate this point.

It would thus appear that in rat liver the proteins are synthesized by ribosomes of similar structural characteristics present in the various intracellular fractions.

Acknowledgment

We are indebted to Severo Ochoa for his guidance and interest.

References

1. SIEKEVITZ, P. 1952. J. Biol. Chem. **195:** 549.
2. LITTLEFIELD, J. W., E. B. KELLER, J. GROSS & P. C. ZAMECNIK. 1955. J. Biol. Chem. **217:** 111.
3. ALLFREY, V. G., A. E. MIRSKY & S. OSAWA. 1957. J. Gen. Physiol. **40:** 451.

4. Logan, R., A. Ficq & M. Errera. 1959. Biochim. et Biophys. Acta. **31:** 402.
5. McLean, J. R., G. L. Cohn, I. K. Brandt & M. V. Simpson. 1958. J. Biol. Chem. **233:** 657.
6. Greengard, O. & P. N. Campbell. 1959. Biochem. J. **71:** 148.
7. Rendi, R. & T. Hultin. Exptl. Cell Research. In press.
8. Rendi, R. 1959. Exptl. Cell Research. **17:** 585.
9. Rendi, R. Exptl. Cell Research. In press.
10. Hoagland, M. B., M. L. Stephenson, J. F. Scott, L. I. Hecht & P. C. Zamecnik. 1958. J. Biol. Chem. **231:** 241.
11. Hall, B. D. & P. Doty. 1959. J. Molecular Biol. **1:** 111.
12. Gierer, A. 1958. Z. Naturforsch. **13b:** 778.
13. Hopkins, J. W. 1959. Proc. Natl. Acad. Sci. **40:** 1461.

SEPARATION OF AMINO ACID-SPECIFIC "SOLUBLE"-FRACTION RIBONUCLEIC ACIDS*

Robert W. Holley, Jean Apgar, B. P. Doctor

United States Plant, Soil, and Nutrition Laboratory, Agricultural Research Service, United States Department of Agriculture, and Department of Biochemistry and Nutrition, Cornell University, Ithaca, N.Y.

If a tissue homogenate or an extract of cells is centrifuged at top speed in the preparative ultracentrifuge, the supernatant or "soluble" fraction is found to contain a small amount of ribonucleic acid (RNA). This RNA, called soluble-RNA, is heterogeneous and is apparently a mixture of at least 20 different RNAs, each of which accepts a specific amino acid from the corresponding amino acid-activating enzyme. The reactions involved are shown in the accompanying equations, which are believed to describe the initial steps in the biosynthesis of proteins.[1-6]

$$\text{Enzyme}_{1\text{-}20} + \text{Amino Acid}_{1\text{-}20} + \text{Adenosine Triphosphate} \rightleftharpoons$$

$$\text{Enzyme}_{1\text{-}20}\text{-Amino Acyl}_{1\text{-}20}\text{-Adenosine Monophosphate} + \text{Pyrophosphate}$$

$$\text{Enzyme}_{1\text{-}20}\text{-Amino Acyl}_{1\text{-}20}\text{-Adenosine Monophosphate} + \text{RNA}_{1\text{-}20} \rightleftharpoons$$

$$\text{Enzyme}_{1\text{-}20} + \text{Amino Acyl}_{1\text{-}20}\text{-RNA}_{1\text{-}20} + \text{Adenosine Monophosphate}$$

The function of the soluble-fraction RNAs is believed to be the transfer of specific amino acids to their sites in the "template" for protein synthesis, and therefore these RNAs are also called transfer-RNAs.[7]

The soluble-fraction or transfer-RNAs furnish the chemist with extremely interesting materials for fractionation and chemical studies. They are key intermediates in the biosynthesis of proteins, which is sufficient reason to study their chemistry. Also, they are amino acid-specific, and it is likely that studies of their structures will furnish clues for the coding problem—the problem of how nucleic acid structure controls protein structure. In addition, these RNAs are of relatively low molecular weight, and they furnish the chemist with his first opportunity to determine the complete structure of a biologically active RNA.

Pure RNAs are required for most chemical studies, for example, for studies of structure. Therefore, the initial problem confronting the chemist is the separation of the different amino acid-specific RNAs, the topic to be discussed in this paper.

There are two basically different approaches to the problem of separation of the different amino acid-specific RNAs. One approach is to apply fractionation methods directly to the mixture of RNAs, hoping to separate the different amino acid-specific RNAs on the basis of differences in the properties of the RNAs themselves. The other approach is to first attach an amino acid to its specific RNA and then exploit the differences in properties that result from this attachment.

* This work was supported in part by Grant G-4861 from the National Science Foundation, Washington, D.C.

At present the meager information available about the structures of the soluble-fraction RNAs indicates similarities in the structures of the different amino acid-specific RNAs rather than differences. All apparently have a terminal adenylic acid group, and it is to the free 2′- or 3′-hydroxyl of this terminal group that the amino acid becomes attached.[8] All apparently have two cytidylic acid units adjacent to the terminal adenylic acid.[9] From theory it may be said that the only difference that would be required in order to have 20 different RNAs would be variations in the sequence of three nucleotides. If this were the only difference between the different RNAs, which probably have molecular weights in the range of 20,000 to 30,000, one would be very pessimistic about the possibility of separating the different RNAs on the basis of differences in properties of the RNAs themselves. The results of countercurrent distribution (CCD) studies to be presented below indicate that such pessimism is unwarranted and, in fact, suggest that there are major structural differences between certain of the amino acid-specific RNAs.

Although this paper is primarily concerned with CCD studies, the results of other fractionation studies should be mentioned briefly. Work is under way in several laboratories using the second approach mentioned above. Of the work that has been mentioned in print, the most promising is that of Geoffrey Brown and his associates. These workers have used the coupling of a diazotized polyaminostyrene with tyrosine as a means of selectively isolating the tyrosine-specific RNA. Their results, presented in preliminary form[10] at the Brookhaven Symposium last June, are very promising.

Of the fractionation methods we have studied, electrophoresis was the first method investigated. Our results were not promising, but the experiments were cursory. Fractionation by means of electrophoresis has been studied much more extensively by Lipmann and his associates.[11]

Chromatography on ECTEOLA-cellulose has also been investigated. ECTEOLA-cellulose has been applied to the fractionation of DNA by Bendich and his associates[12,13] and to the fractionation of RNA by Bradley and Rich.[14] In our experiments, using soluble-fraction RNA prepared from rat liver, a series of ultraviolet-absorbing peaks were obtained, and there seemed to be some fractionation, based on activity. However, rechromatography of a single peak gave a series of peaks. This type of behavior had been observed by Bradley and Rich in their studies of the fractionation of other RNAs. Zamecnik and his associates have also studied the chromatography of soluble-fraction RNA on ECTEOLA-cellulose.[15]

Fractionation by chromatography on Cato-8 starch has been reported by Smith *et al.*,[7] who observed relative enrichment of the tyrosine-specific RNA in eluates of lower salt concentration and of the leucine-specific RNA in eluates of higher salt concentration. These workers also report evidence for separation by ammonium sulfate fractionation, and Schweet *et al.*[16] have mentioned preliminary studies using calcium phosphate gel.

Two years ago R. C. Warner and P. Vaimberg described a solvent system for CCD of RNAs.[17] The solvent system was composed of 2 *M* *p*H 8 phosphate buffer, isopropanol, and formamide. These workers reported CCD of commercial yeast RNA and of RNA prepared from *Azotobacter vinelandii*. Using their solvent system, we found that soluble-fraction RNA prepared from

rat liver gave a gross partition coefficient close to one. Distribution of the rat-liver RNA resulted in a broad distribution curve, as would be expected for heterogeneous material. Redistribution of different fractions obtained from the first distribution gave peaks at tubes corresponding to the different starting fractions, indicating that CCD was giving real fractionation of the rat-liver RNAs.[18]

The fractionation of the RNAs observed could, of course, have been due to separation of inactive or degraded RNA from the active RNAs rather than to separation of different amino acid-specific RNAs. These possibilities were distinguished by assays of the RNA fractions after reisolation of the fractions in form suitable for assay. The assays established that amino acid-specific RNAs were being separated.[19] Of particular interest also was the finding that each of the amino acid-specific RNAs distributed as a single peak, consistent with the assumption that only one RNA is involved as an acceptor for each amino acid. The separation of certain of the amino acid-specific RNAs was quite promising. For example, after 250 transfers, the peaks for alanine and tyrosine activity were separated by approximately 25 tubes. Other activities were not separated, for example, alanine and threonine.

We were interested in extending the separation as far as possible and were very fortunate at this stage to obtain the collaboration of L. C. Craig and T. P. King of the Rockefeller Institute. The results of an 1100-transfer distribution carried out with their collaboration in their 1000-tube instrument are shown in FIGURE 1. The heterogeneity of the rat-liver RNA is apparent from the series of ultraviolet-absorbing peaks. The partition coefficient (K) was observed to increase gradually across the tubes, as would be expected for very heterogeneous material. The distribution shown in FIGURE 1 was continued to a total of 3000 transfers, the effluent from the 1000th tube being collected in an automatic fraction collector. Some of the fractions were lost and data are incomplete, but sufficient data were obtained to indicate that certain of the amino acid-specific RNAs, for example the alanine and threonine RNAs, were still only partially separated. It was also found that losses of activity, which had been observed in the earlier distributions,[19] were serious during the extended distribution and reisolation procedures.

Actually, the loss of activity was not entirely unexpected. The solvent system of Warner and Vaimberg[17] has a *p*H of approximately 8, which is high enough to cause one to expect slow hydrolysis of RNA. It seemed desirable, therefore, to attempt to develop a solvent system of lower *p*H. Such a system was developed by using 1.9 *M* *p*H 6 phosphate buffer, and a preliminary distribution in this system indicated that the separation of different amino acid-specific RNAs was actually improved at the lower *p*H. This improved separation was unexpected since Warner and Vaimberg suggested [17] from their results that separation of RNA fractions in the *p*H 8 system was due to variations in the content of guanine and uracil in which the phenolic groups would be partially ionized at *p*H 8. The observation that separation of the different amino acid-specific RNAs was improved at *p*H 6 suggested that variations in guanine and uracil content was not the basis for the separation observed in our distributions.

Having found the improved separation in the *p*H 6 system and anticipating

much greater stability in this system, we again enlisted the collaboration of Craig and King, used 150 mg. of rat-liver "soluble"-fraction RNA, and carried out a distribution in the 1000-tube instrument. To our dismay, the reisolated RNA fractions were essentially inactive. We still have little evidence as to the cause of this loss of activity. A sufficient quantity of the particular RNA sample used in the distribution was left to establish that it was unstable to shaking in the *p*H 6 system, but the RNA sample was soon exhausted and other samples were found to be more stable. We have some indication, using our 200-tube instrument, that the addition of magnesium chloride to the countercurrent system has a stabilizing influence.

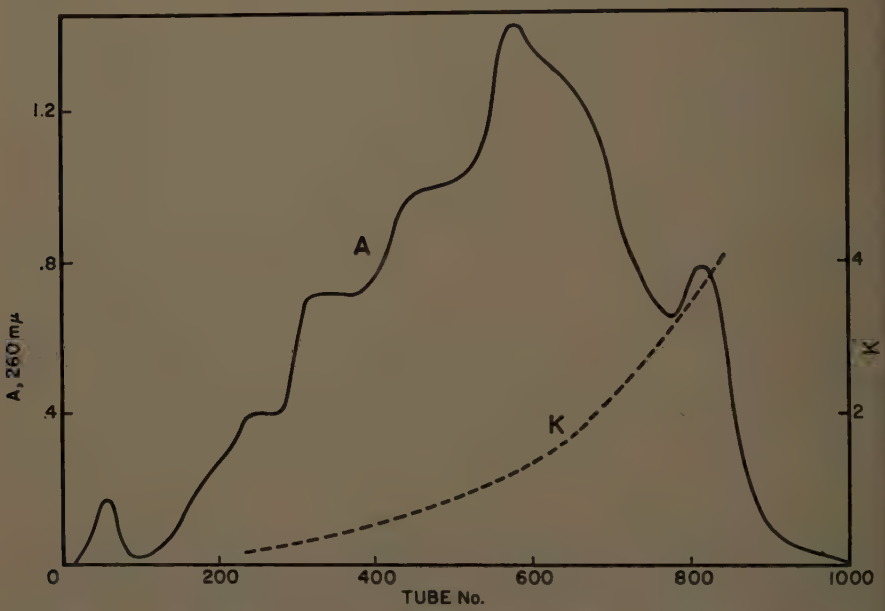


FIGURE 1. 1100-Transfer CCD of rat-liver "soluble" fraction RNA in *p*H 8 system: solid line, absorbency (A) at 260 mμ; broken line, partition coefficient (K).*

In recent CCD experiments, magnesium chloride has been added to the *p*H 6 system, and certain modifications have been made in the procedure used for reisolation of the RNA fractions for assay: specifically, dialysis against dilute magnesium chloride solution and evaporation of the dialyzed solution instead of lyophylization. The recovery of activity has been approximately 80 per cent from 200 transfer distributions. FIGURE 2 shows the results of a distribution of rat-liver "soluble"-fraction RNA in the *p*H 6 system. The peaks of activity for threonine and tyrosine are separated by approximately 40 tubes after 200 transfers, as compared to separation of approximately 25 tubes after 250 transfers in the *p*H 8 system. FIGURE 3 shows a 200-transfer distribution of a preparation of amino acid-specific RNAs from yeast. This

* In collaboration with L. C. Craig and T. P. King, The Rockefeller Institute, New York, N.Y.

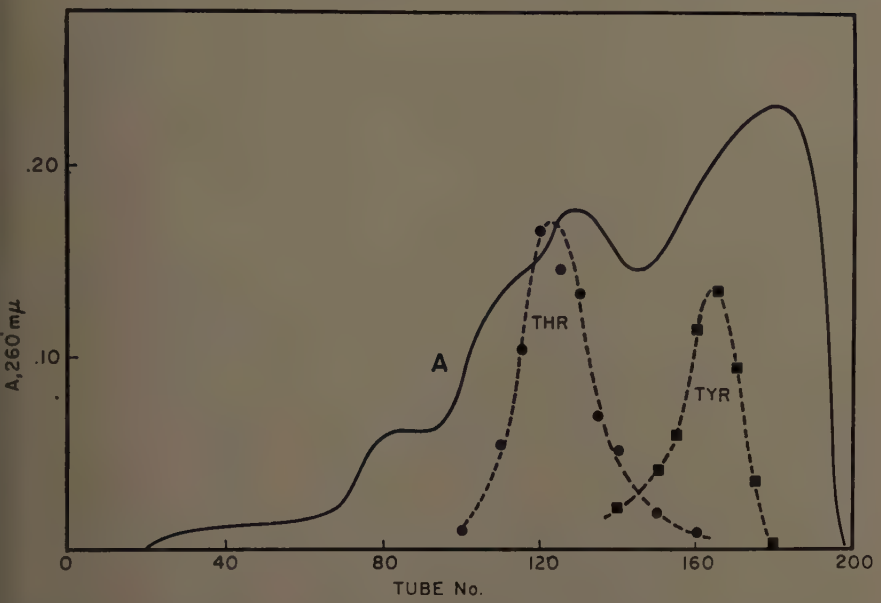


FIGURE 2. 200-Transfer CCD of rat-liver "soluble" fraction RNA in ϕH 6 system solid line, absorbency (A) at 260 mμ; dots in broken line, threonine-acceptor activity squares in broken line, tyrosine-acceptor activity.

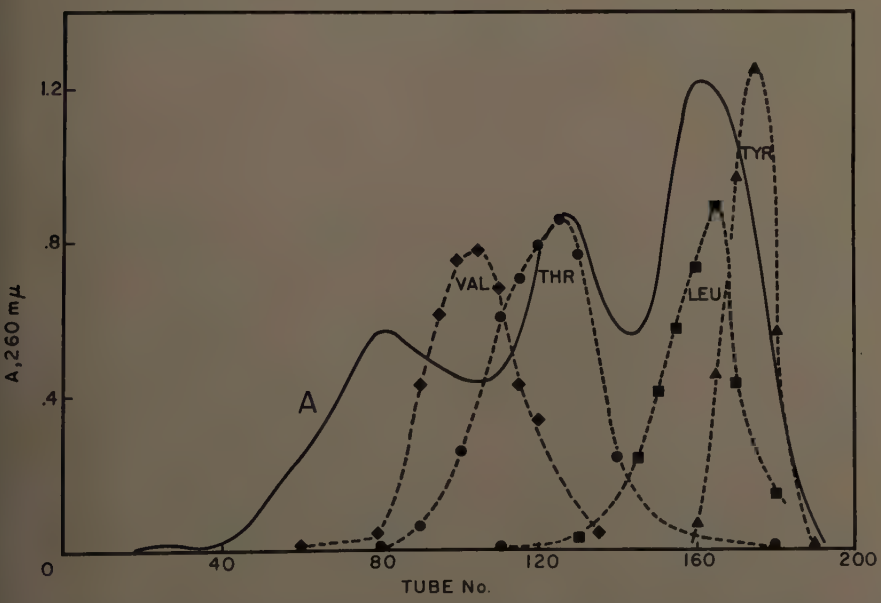


FIGURE 3. 200-Transfer CCD of yeast amino acid-acceptor RNA in ϕH 6 system: solid line, absorbency (A) at 260 mμ; diamonds in broken line, valine-acceptor activity; dots in broken line, threonine-acceptor activity; squares in broken line, leucine-acceptor activity; triangles in broken line, tyrosine-acceptor activity.

RNA was prepared directly from bakers' yeast by a modification of the method of Monier *et al.*[20] (we are indebted to R. Monier for making this procedure available to us prior to publication). The great advantage of the yeast RNA is its ready availability. As may be seen in FIGURE 3, the threonine- and tyrosine-specific RNAs are again almost completely separated after 200 transfers. The positions of the leucine- and valine-RNAs are also shown. The tyrosine-RNA fraction recovered from this distribution had a specific activity six times that of the starting material.

It is of great interest that the threonine- and tyrosine-acceptor RNAs show essentially the same distribution behavior whether isolated from rat liver or bakers' yeast. This suggests that the differences in the structures of the threonine- and tyrosine-RNAs that lead to the widely different partition coefficients are common to the RNAs isolated from these widely different sources.

The observed marked difference in the partition coefficients of the threonine- and tyrosine-RNAs (and also of the valine- and leucine-RNAs) is one of the most interesting points to come out of these studies. The difference is so great that almost complete separation of these pairs of RNAs can be achieved using a few separatory funnels. It seems most unlikely that such a difference in partition coefficient is caused by a minor structural difference such as variation in the sequence of a few nucleotides. In fact, it is likely that there are major structural differences between the different RNAs. Whatever the interpretation, it is clear from these results that different RNAs can be separated on the basis of differences in the properties of the RNAs themselves—knowledge that will be very important in encouraging the development of new methods for the fractionation of these RNAs.

Experiments are under way to establish the nature of the structural differences that lead to such widely different partition coefficients.

Experiments are also being conducted to improve and extend the CCD procedures.

References

1. GLASSMAN, E., E. H. ALLEN & R. S. SCHWEET. 1958. J. Am. Chem. Soc. **80:** 4427.
2. HOLLEY, R. W. & J. GOLDSTEIN. 1959. J. Biol. Chem. **234:** 1765.
3. MAGER, J. & F. LIPMANN. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 305.
4. OGATA, K. & H. NOHARA. 1957. Biochim. et Biophys. Acta. **25:** 659.
5. PREISS, J., P. BERG, E. J. OFENGAND, F. H. BERGMANN & M. DIECKMANN. 1959. Proc. Natl. Acad. Sci. U. S. **45:** 319.
6. ZAMECNIK, P. C., M. L. STEPHENSON & L. I. HECHT. 1958. *Ibid.* **44:** 73.
7. SMITH, K. C., E. CORDES & R. S. SCHWEET. 1959. Biochim. et. Biophys. Acta. **33:** 286.
8. ZACHAU, H. G., G. ACS & F. LIPMANN. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 885.
9. HECHT, L. I., M. L. STEPHENSON & P. C. ZAMECNIK. 1958. Biochim. et Biophys. Acta. **29:** 460.
10. BROWN, G. L., A. V. W. BROWN & J. GORDON. 1959. Brookhaven Symposia in Biol. No. 12 (BNL558; C-29). : 47.
11. LIPMANN, F., W. C. HÜLSMANN, G. HARTMANN, H. G. BOMAN & G. ACS. 1959. J. Cellular Comp. Physiol. **54**(supp. 1)**:** 75.
12. BENDICH, A., J. R. FRESCO, H. S. ROSENKRANZ & S. M. BEISER. 1955. J. Am. Chem. Soc. **77:** 3671.
13. BENDICH, A., H. B. PAHL, G. C. KORNGOLD, H. S. ROSENKRANZ & J. R. FRESCO. 1958. *Ibid.* **80:** 3949.
14. BRADLEY, D. F. & A. RICH. 1956. *Ibid.* **78:** 5898.
15. HOAGLAND, M. B., M. L. STEPHENSON, J. F. SCOTT, L. I. HECHT & P. C. ZAMECNIK. 1958. J. Biol. Chem. **231:** 241.

16. Schweet, R. S., F. C. Bovard, E. Allen & E. Glassman. 1958. Proc. Natl. Acad. Sci. U. S. **44:** 173.
17. Warner, R. C. & P. C. Vaimberg. 1958. Federation Proc. **17:** 331.
18. Holley, R. W. & S. H. Merrill. 1959. J. Am. Chem. Soc. **81:** 753.
19. Holley, R. W., B. P. Doctor, S. H. Merrill & F. M. Saad. 1959. Biochim. et Biophys. Acta. **35:** 272.
20. Monier, R., M. L. Stephenson & P. C. Zamecnik. Biochim. et Biophys. Acta. In press.

SYNTHETIC ASPECTS OF RIBOSOMES

Richard B. Roberts
Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, D. C.

Introduction

The methods for investigating the reaction sequences of a cell fall into two general classes. In one the cells are disrupted and fragments are isolated that catalyze one particular reaction. This method has been successful in isolating the enzymes that accomplish the synthesis of small molecules; more recently enzymes capable of polymerizing RNA and DNA have been purified. Applied to the study of protein synthesis, this technique has focused interest on the occurrence of amino acid-activating enzymes and the linkage of amino acids to soluble RNA. In demonstrating actual systhesis of protein there has been less striking success. The quantity and rate of protein synthesis carried out by fragments of disrupted cells are disappointingly low. Increased specific activity of the preparation, usually the criterion for the successful isolation of an enzyme, has not been achieved. This may be an indication that protein synthesis is carried out by larger structures that must remain intact.

The alternative method is to observe reactions as they occur in intact growing cells. Deficient mutants, isotopic competition, and kinetic measurements of tracer-labeled materials have been used to study the reactions among the smaller molecules. In general the results confirm the reaction schemes based on studies of isolated enzymes. In addition these techniques often can indicate the flow of material along different pathways.[1] Such a confirmation is essential because the isolation of an enzyme from a cell does not guarantee that it was active in the cell.

At present the evidence from studies of intact cells on the role of the activating enzymes is not conclusive.[2] Furthermore, the kinetics of the soluble RNA-amino acid complex do not seem (to me at least) to follow the course expected if the S-RNA-amino acids were compulsory intermediates in protein synthesis. The amino acid-lipid complex observed by Hendler[3] seems to be an equally likely candidate as a carrier in protein synthesis. Alternatively, it may prove (R. W. Hendler, unpublished observations) to be the carrier[4] required in the entry of amino acids into the cell.

However, these differences between the whole cell and disrupted cell experiments should not be overemphasized. It is always easy to miss distinctions, particularly when dealing with small quantities of highly labile material. The positive results obtained by observing the synthesis of protein and ribosomes in intact, growing *Escherichia coli* are more significant. Because of the time limitations, only the conclusions and a few of the experiments that lead to these conclusions can be presented.

Properties of Ribosomes

FIGURE 1 shows the picture obtained by disrupting a growing culture of *E. coli* and examining the juice in the analytical centrifuge. There are clearly a number of distinct groups of particles indicated by their sedimentation con-

stants. These numbers are rounded off to indicate that they are for identification only and not to indicate precise values. The larger particles can be interconverted by changing the magnesium concentration; the particle labeled 85S has a sedimentation constant that varies continuously between 70S and 100S depending on the magnesium concentration.[5] Evidently the difference between the 70S and 85S particle is one of shape or hydration. Furthermore,

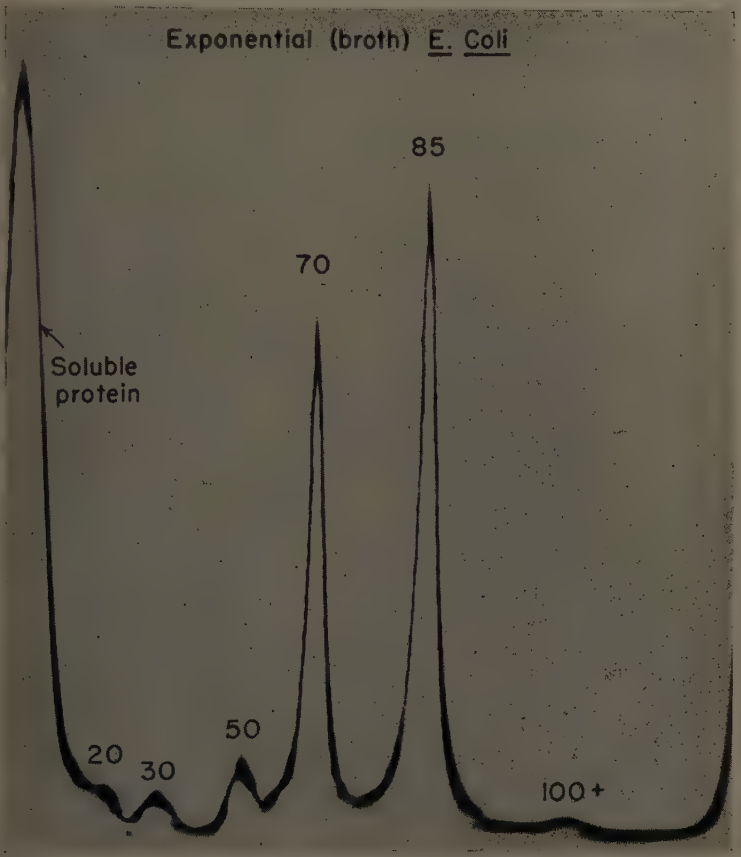


FIGURE 1. Ribosome pattern of *E. coli* growing exponentially in broth medium as shown in the analytical centrifuge.

as the magnesium concentration is reduced even further, the 70S particle breaks down into 30S and 50S particles.[6,7] Nevertheless, these different groups are not simply an equilibrium mixture—the proportions have biological significance. For example, the 85S particles accumulate rapidly when ribosome synthesis is stopped by lack of an energy source.[8] The 70S particles are the most active in protein synthesis, and the smaller particles are the first to show incorporation of tracers into the structure of the ribosomes.

The 30S and 50S particles present in the cell juices are distinct from the 30S and 50S particles that result from the breakdown of the 70S particle. The

20S group is actually made up of two classes: (1) slightly heavier particles (~22S) that are rich in protein and (2) lighter particles (~18S) that contain nucleic acid.

With the exception of the 20S group, these particles have been separated from each other and from contaminating protein and nucleic acid by the sedimentation of a layer through a sucrose gradient,[9] and their compositions have

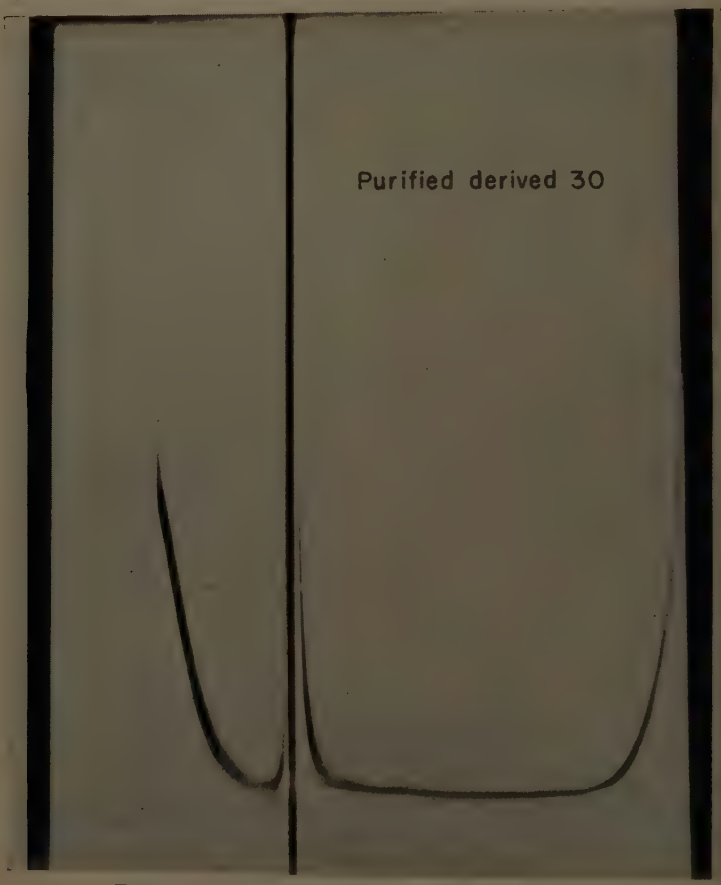


FIGURE 2. 30S ribosomes purified by sedimentation.

been measured. FIGURE 2 shows the purity that can be achieved. These purified particles all have protein-nucleic acid ratios that correspond to two amino acids per nucleotide,[10] indicating that their basic structure may be three-stranded. The base ratios of the RNA of the 30S and 50S particles are slightly different, the 50S group having a greater excess of purine[11] (TABLE 1).

Synthesis of Protein by Ribosomes

To show the role of the bacterial ribosomes in synthesizing protein,[12] the distribution of a tracer material is observed after a very short period of incorporation terminated by sudden chilling. FIGURE 3 shows that $S^{35}O_4^=$ is a very

TABLE 1
COMPOSITION OF *E. COLI* RIBONUCLEIC ACIDS

	Soluble		Particle				
	S	CA-B	30(n)	30(70)	50(n)	50(70)	CA-D
C	29.1	27.2	22.2	23.6	21.0	20.5	20.2
A	19.7	19.9	24.2	24.2	26.4	26.4	28.6
G	34.2	35.6	30.4	31.6	34.1	34.8	32.2
U	17.2	17.7	(23.1)	20.5	18.5	18.3	19.0

Values are the arithmetic means of several determinations for each component. The uncertainty is ±3 per cent of the given value except for the bracketed value where the uncertainty was ±7 per cent. 30(n) and 50(n) indicate 30S and 50S particles present in cell juice. 30(70) and 50(70) indicate 30S and 50S particles derived from 70S particles. CA-B and CA-D indicate soluble and particle type RNA synthesized in presence of chloramphenicol, separation by chromatography on DEAE.

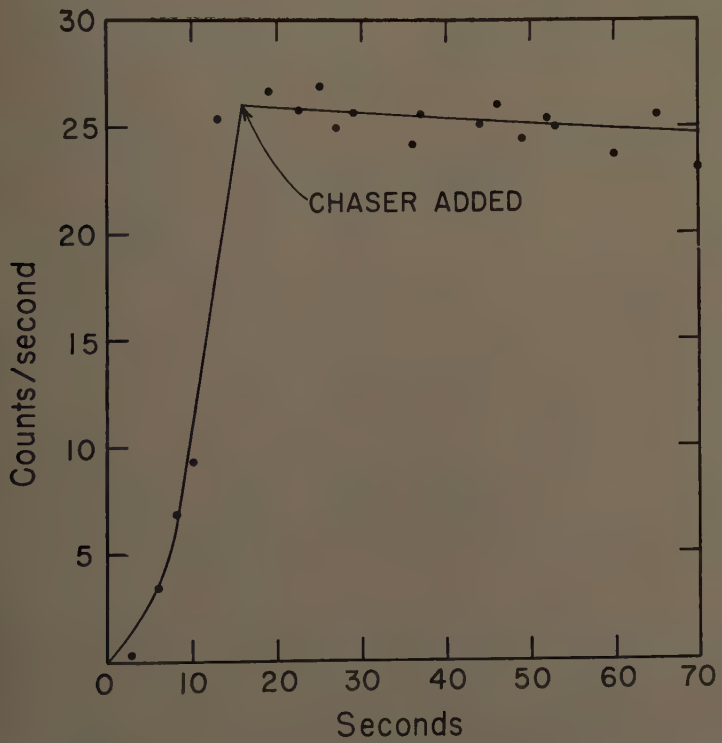


FIGURE 3. Time course of incorporation. $S^{35}O_4^{=}$ was added to a growing culture of sulfur-depleted *E. coli* at time 0. Samples were withdrawn at indicated times and squirted into TCA, filtered, and counted. At 16 sec. $S^{32}O_4^{=}$, S^{32} cystine, and S^{32} methionine were added. Note prompt incorporation of the tracer into TCA-precipitable material and rapid cessation of incorporation after addition of nonradioactive material.

FIGURE 4. Sedimentation analysis of cell juice after 15 sec. incorporation of $S^{35}O_4^=$. $S^{35}O_4^=$ was added to a growing culture of sulfur-depleted *E. coli*. After 15 sec. the culture was chilled, harvested, washed, and broken. The cell walls and membranes were spun out, and the juice was analyzed by sedimentation in the swinging-bucket rotor. (*a*) Cells incubated 15 sec. with $S^{35}O_4^=$. (*b*) Cells incubated 15 sec. with $S^{35}O_4^=$ followed by 15 sec. incubation with S^{32} "chaser." (*c*) Fifteen sec. incubation with $S^{35}O_4^=$ followed by 120 sec. with S^{32} "chaser." Note transfer of radioactivity from 70-85S region to nonsedimenting region. Centrifugation 75 min. at 37,000 rpm.

suitable tracer. Sulfur-depleted cells show a very prompt incorporation of $S^{35}O_4^=$ into TCA-precipitable material and an equally prompt cessation of the incorporation when the tracer is diluted by a nonradioactive "chaser." FIGURE 4 shows the distribution of newly incorporated sulfur among the particles and the soluble proteins. A large proportion of the S^{35} sediments with the particles.

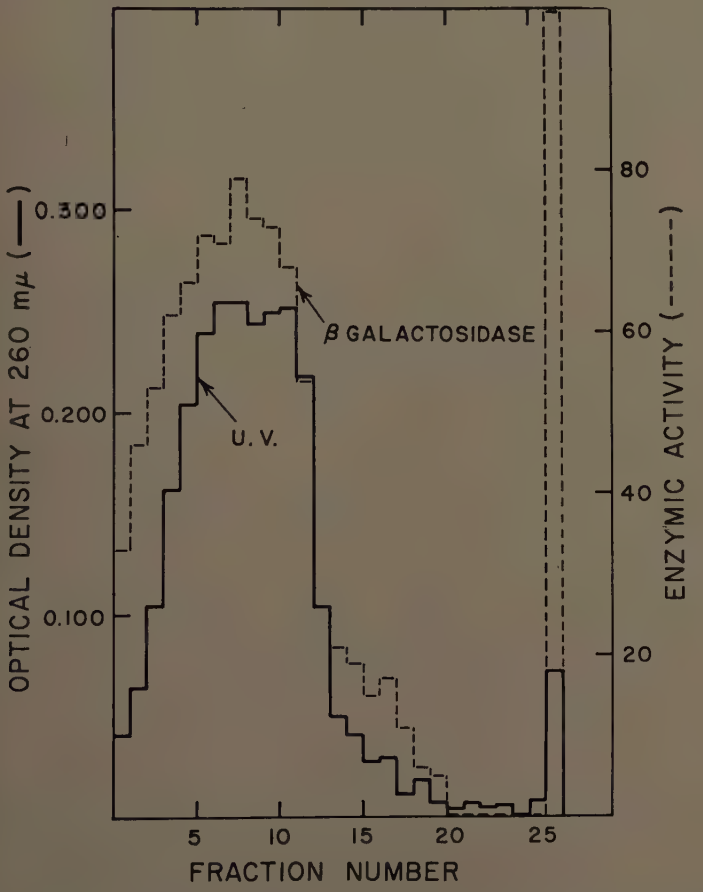


FIGURE 5. Sedimentation analysis of purified particles shows β-galactosidase activity corresponding to UV absorption of particles.

A quantitative transfer from the particles to the soluble fraction occurs after addition of the chaser. The actual quantity of nascent protein corresponds roughly to the quantity of protein synthesized during 3 sec.

The sedimentation constant of the nascent protein corresponds to that of the ribosomes. It would not be expected that the presence of a polypeptide strand of 200 amino acids would influence the sedimentation constant appreciably. Were each of these 200 amino acids linked to the ribosomes through a strand of soluble RNA, the sedimentation constant would be considerably altered.

Another indication of the role of the ribosomes in synthesis is the association of small quantities of enzymes (presumably nascent) with the ribosomes.[13,14] FIGURE 5 shows the correspondence of β-galactosidase activity with the UV absorption of the particles. The enzyme of the particle fraction has the interesting property that its activity is increased appreciably by treatment with its antibody. Roughly 0.5 to 1 per cent of the enzymic activity of uninduced cells is associated with the particles.

Synthesis of Ribosomes

$P^{32}O_4^{\equiv}$ provides a useful tracer to observe the synthesis of the nucleic acid portion of the ribosomes.[10,15] Unfortunately, this tracer must pass through a

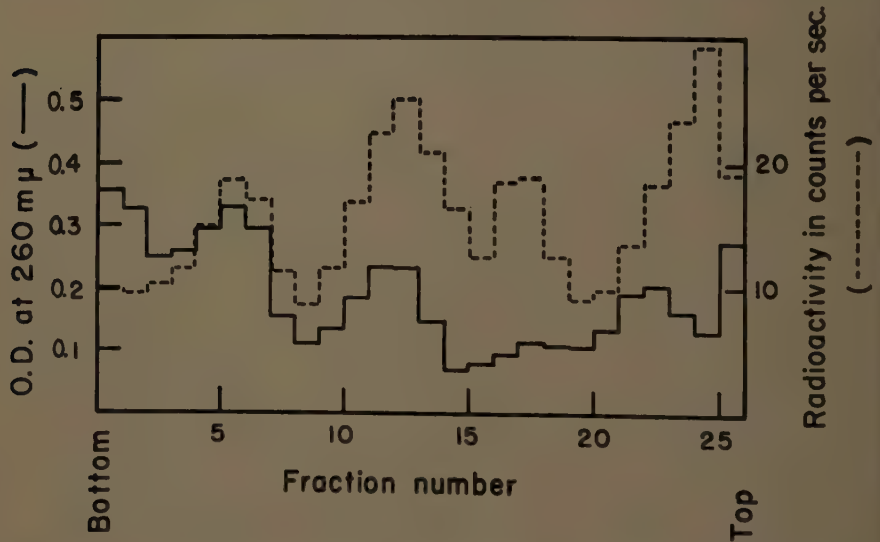


FIGURE 6. Sedimentation analysis of ribosomes from *E. coli* labeled by uptake of P^{32} for 2 min.

large metabolic pool that introduces kinetic delays. The different groups of ribosomes (FIGURE 6) show further kinetic delays in addition to those introduced by the nucleotide pool. At early times, the small particles have considerably higher specific radioactivities than do the large ones as listed in TABLE 2. A kinetic sequence in which P^{32} passes from the small particles to larger ones is clearly indicated. Similar product-precursor relationships are indicated by chromatographic analysis on DEAE. FIGURE 7 shows the bulk of the ultraviolet material in the region where the large particles elute and the bulk of the newly incorporated P^{32} in the region where the small particles appear. Electrophoresis will not resolve the two groups.

Cells given a "persistent" label by growing them in a radioactive medium followed by several generations' growth in a nonradioactive medium show radioactivity in all classes of particles with only slightly higher specific radioactivity in the large ones. Evidently the small particles arise in part by *de*

novo synthesis but mostly from the breakdown of the larger ones. Possibly the rate of breakdown is related to the rate of protein synthesis.

$S^{35}O_4^=$ used as a tracer to show the synthesis of the protein portion of the ribosomes shows essentially the same course of events (FIGURE 8). Chromatography on DEAE reveals one additional feature. One half the protein of

TABLE 2

SPECIFIC RADIOACTIVITIES OF PARTICLES FROM CELLS GIVEN 2′ INCUBATION WITH $P^{32}O_4^=$*

Isolated by sedimentation		Isolated by chromatography	
20S	4.5	Region D	1.9
30S	3.0		
50S	1.6		
30S } From 70S	0.3	Region A	.11
50S } and 85S	0.1		
Soluble RNA	—	Region B	.45

* See FIGURES 6 and 7.

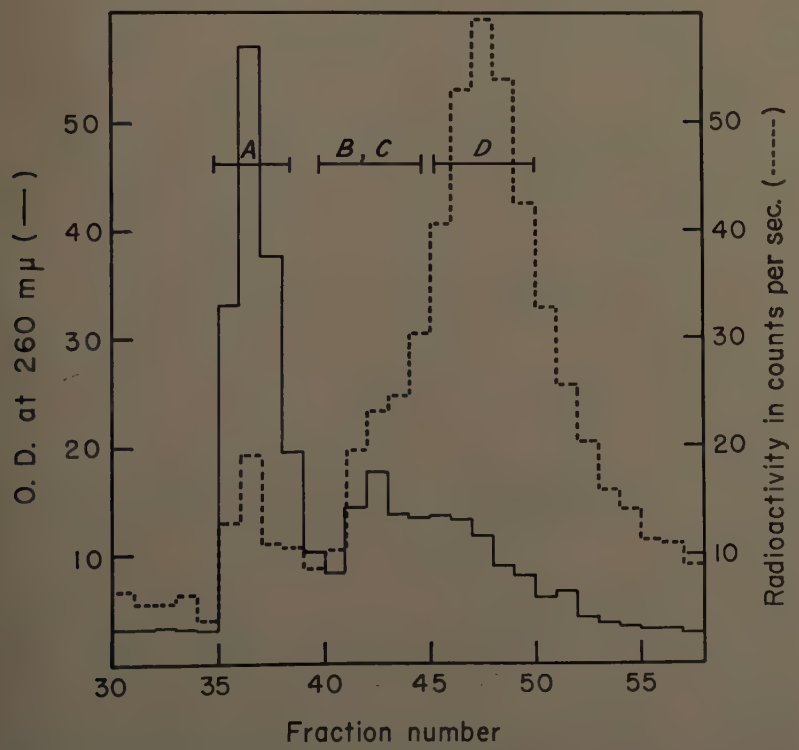


FIGURE 7. DEAE-cellulose chromatography of nucleoproteins and nucleic acids of *E. coli*. Region *A* is nucleoprotein derived from large ribosomes, *B* is soluble RNA, *C* is DNA, and *D* is derived from precursor small ribosomes. The solid line indicates the optical density at 260μ, and the dotted line shows P^{32} that the bacteria had incorporated during a 2-min. labeling period.

ribosomes is lost and does not elute from DEAE. Newly incorporated $S^{35}O_4^=$ is entirely lost, showing that new material is laid down on a pre-existing core.

The results of these and other experiments lead to the model of ribosome synthesis shown in FIGURE 9. This is the only diagram that we have been able to devise that is compatible with the numerous experimental results.

The synthesis of soluble RNA is not included in this diagram because it seems to be an entirely unrelated process; it shows no kinetic relationship to synthesis

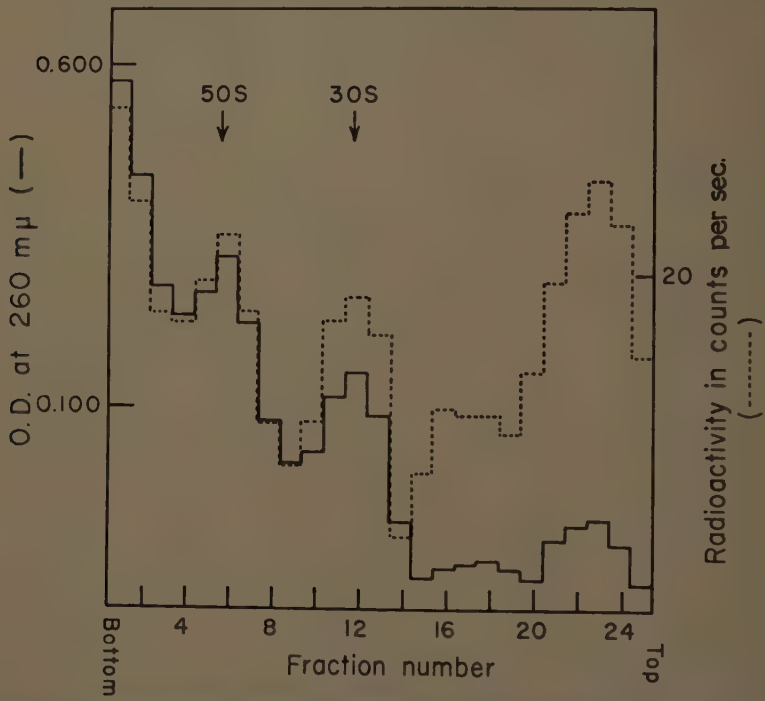


FIGURE 8. Sedimentation analysis of ribosomes from *E. coli* labeled by uptake of S^{35} for 1 min. Reproduced by permission of *Science*.[9]

of the particles. Furthermore, the compositions of the two types of RNA are quite different and there is no exchange of $P^{32}O_4^=$ between them.

Role of DNA

In the experiments mentioned above there is little indication of any role of DNA in the synthesis of proteins or ribosomes. On the contrary, after magnesium starvation, which depletes the content of ribosomes to 5 per cent of normal, the synthesis of ribosomes follows an autocatalytic course.[16]

There is one slight clue to a role of DNA in the formation of the 20S particles. If these particles were formed solely from the splitting of 30S particles, they would have specific radioactivities lower than the 30S group. In fact, the 20S group invariably has the highest specific radioactivity at early times; thus it seems to be synthesized *de novo*. Perhaps DNA serves as template for the

synthesis of 20S particles that then carry the required information from the DNA to the semiautonomous ribosome system.

The work of Caro[17] provides a better indication that DNA may serve as a site for the processes involved in ribosome synthesis. Caro finds that newly incorporated uridine has the same spatial distribution in the cell as does thymidine.

For more direct evidence of the role of DNA we have turned to the 15T⁻A⁻U⁻ mutant of *E. coli*. This cell requires thymine (T), arginine (A), and uracil (U) so that any of the major synthetic systems may be turned off or on, by appropriate choice of supplements to the medium.

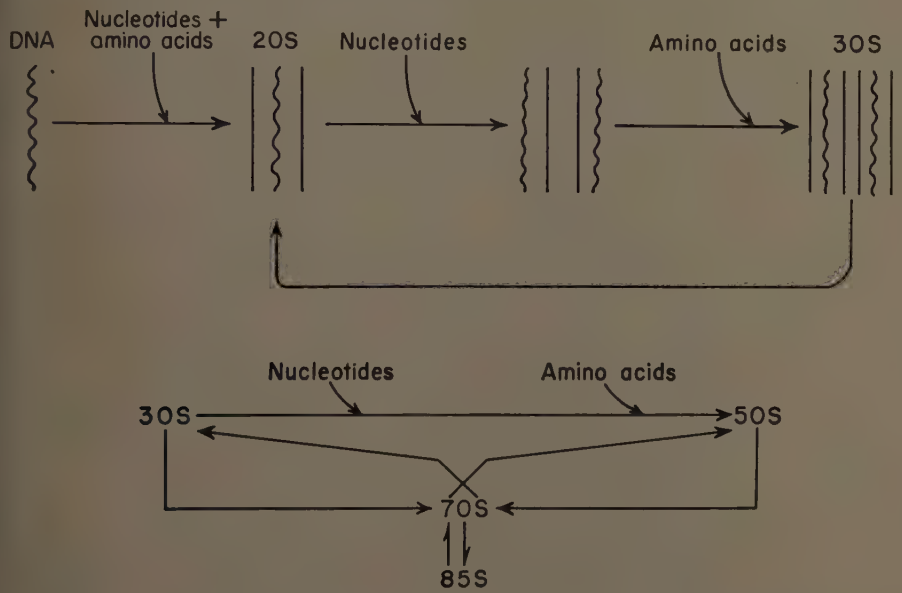


FIGURE 9. Model of ribosome synthesis. DNA may act as template in synthesis of 20S particles.

Cohen has shown that these cells suffer "thymineless death" when taken from a growing culture and incubated without thymine.[18] More recently Maaloe and Hannawalt have found that "thymineless death" can also occur without unbalanced growth when cells are incubated in a medium lacking thymine, arginine, and uracil (personal communication). Accordingly, the loss of viability appears to be caused by an alteration either of the DNA or the nucleus.

We have observed various synthetic processes during the course of incubation without thymine (−T + AU).[19] For the first 30 min. synthesis of the ribosomes and soluble proteins appears normal, and the cells can be induced to form β-galactosidase. Subsequently, as the cells begin to lose viability, the capacity to be induced is lost, abnormal ribosomes accumulate and, finally, protein synthesis stops (FIGURES 10 and 11). Similar losses of synthetic capacities occur when the cells are incubated without thymine, arginine, and

uracil (−T−AU). Intact DNA, but not DNA synthesis, is apparently required for synthesis of ribosomes and the induction of β-galactosidase.

When this mutant is incubated with thymine to permit DNA synthesis, but without arginine or uracil or both, it proceeds to a condition that we call

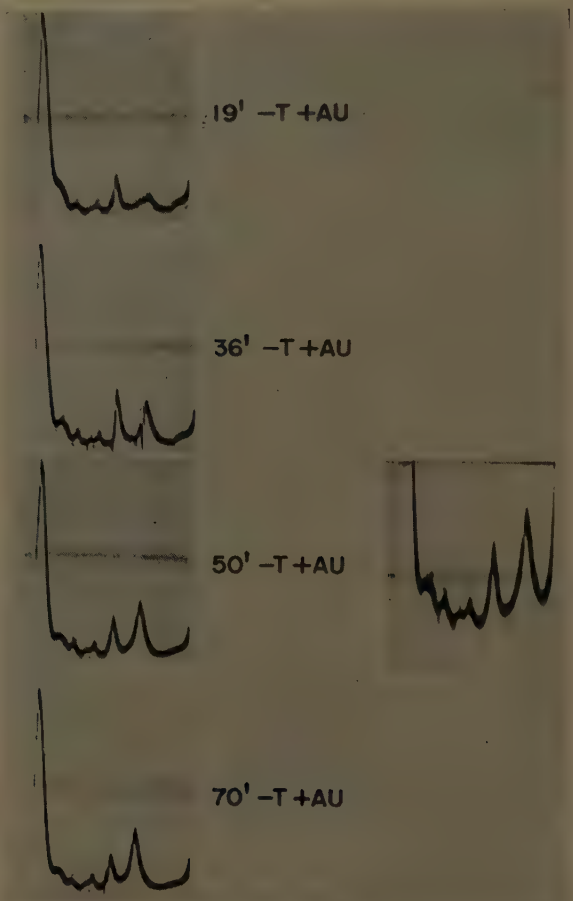


FIGURE 10. Changes in ribosome pattern during growth of 15 TAU⁻ in medium lacking thymine. Accumulation of 85S particles is typical of reduced metabolic rate. Note appearance of new particle group adjacent to 50S group.

thymine-saturated. In the course of this incubation, the viable cell count and the quantity of acid-insoluble thymine increase about 50 per cent (FIGURE 12). Cells brought into the condition have properties that are quite different from exponentially growing cells. Maaloe and Hannawalt (personal communication) noted that thymine-saturated cells do not suffer "thymineless death" when incubated without thymine, arginine, and uracil. In addition, we have found that they are more radiation-resistant and they do not lose their synthetic capacities as readily when incubated without thymine (FIGURE

13). They attributed the immunity to the state of the DNA; that is, that DNA synthesis had gone to completion.

These thymine-saturated cells show a second and perhaps more subtle relationship between the state of the DNA and the induction of β-galactosidase. FIGURE 14 shows that when arginine and uracil are restored the synthesis of

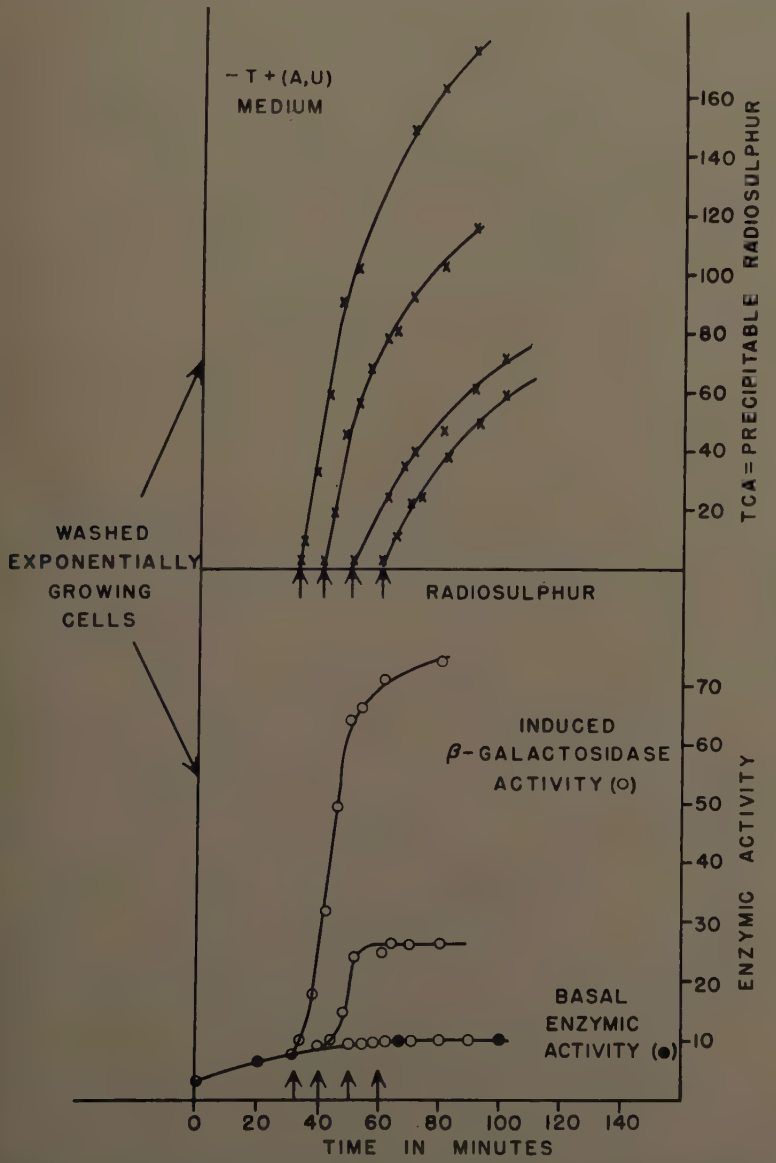


FIGURE 11. The rate of protein synthesis, as measured by S^{35} incorporation, decreases after a 50-min. incubation without thymine. After 50 min. the cells cannot be induced. Arrows indicate times of addition of $S^{35}O_4^{=}$ or inducer.

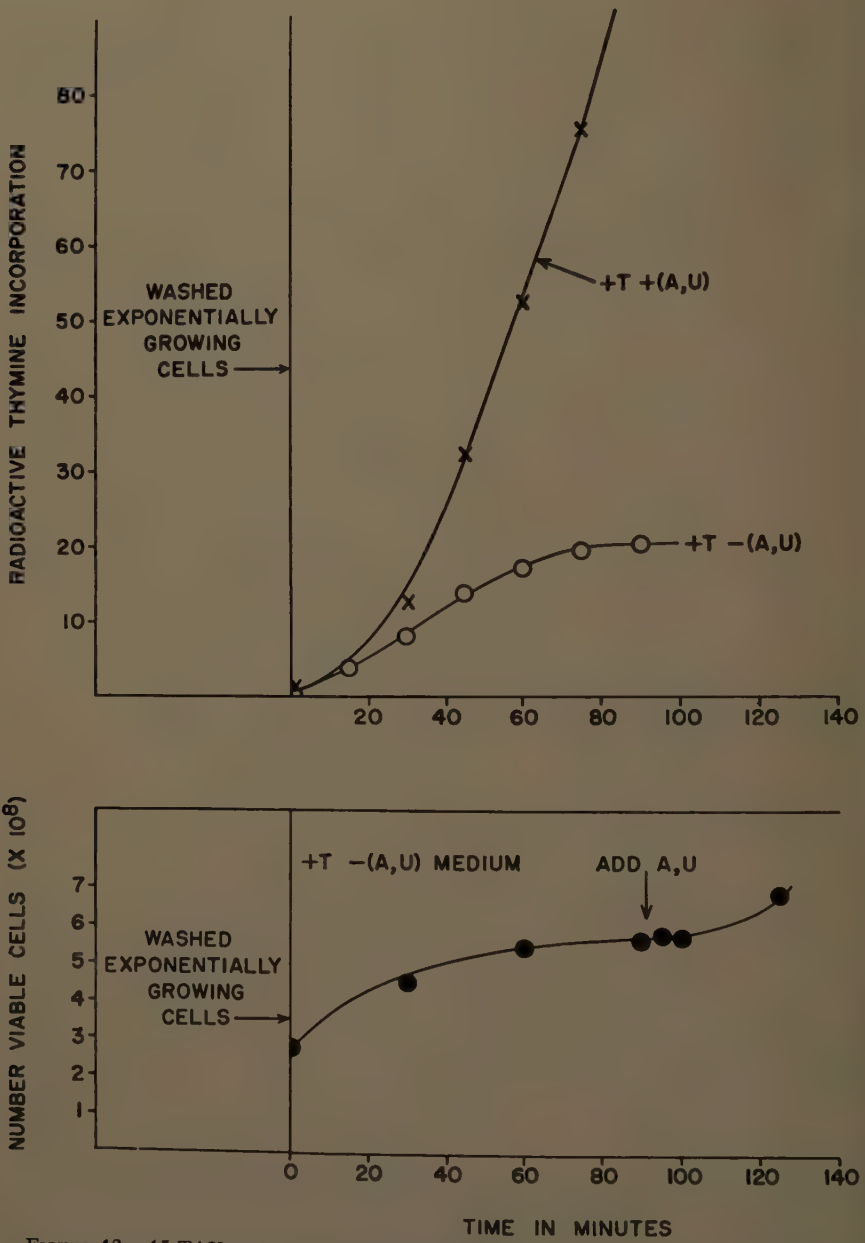


FIGURE 12. 15 TAU⁻ cells continue to incorporate thymine for 90 min. in the absence of arginine and uracil. The cells are then thymine-saturated.

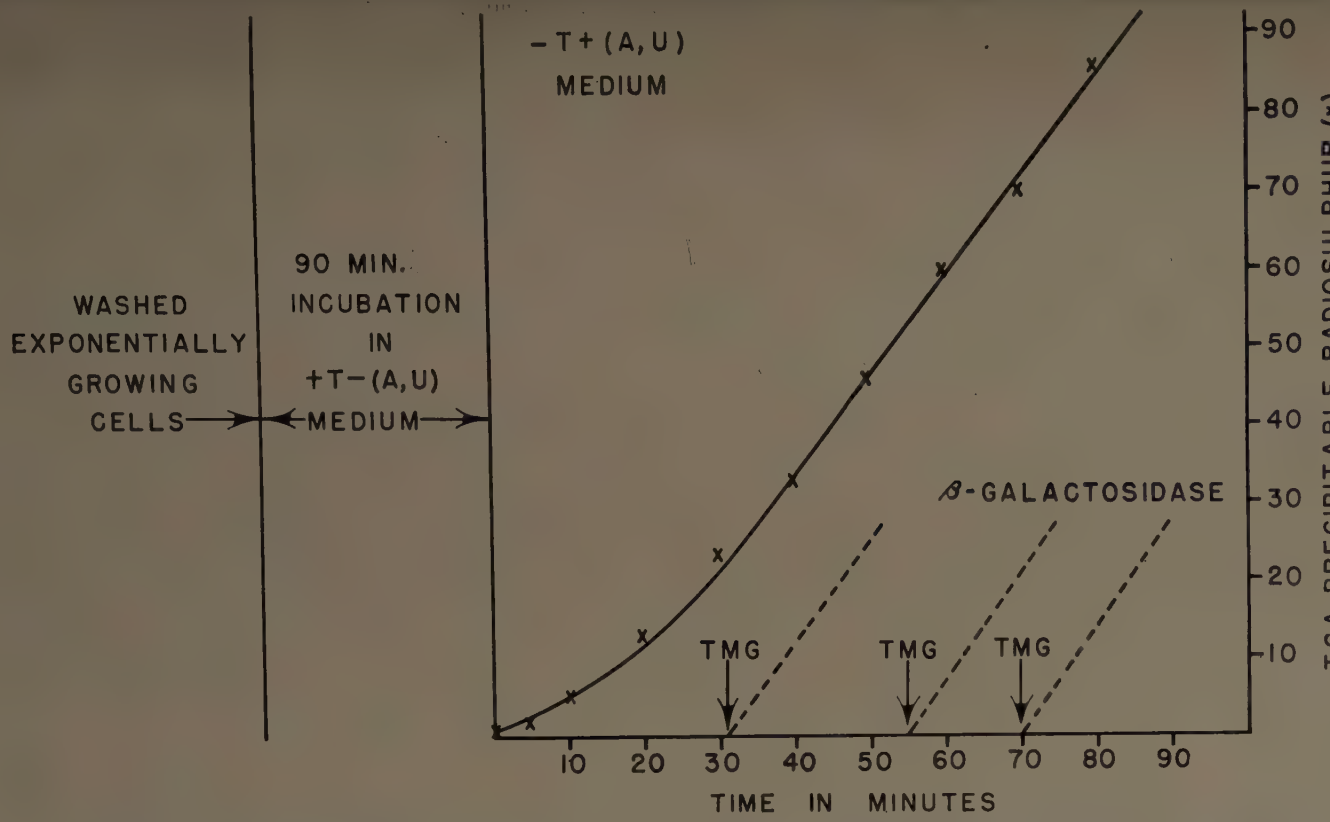


FIGURE 13. Thymine-saturated cells continue to incorporate sulfur and to be inducible after 90 min. incubation without thymine.

RNA and protein is initiated promptly. Chromatographic analysis shows that these proteins are quite normal. Ribosome synthesis during the first 7 min. is also normal. In contrast, thymine incorporation and enzyme induction proceed only after a 12 min. lag. The lag may be removed by restoring arginine and uracil in the absence of the thymine and the inducer (FIGURE 15).

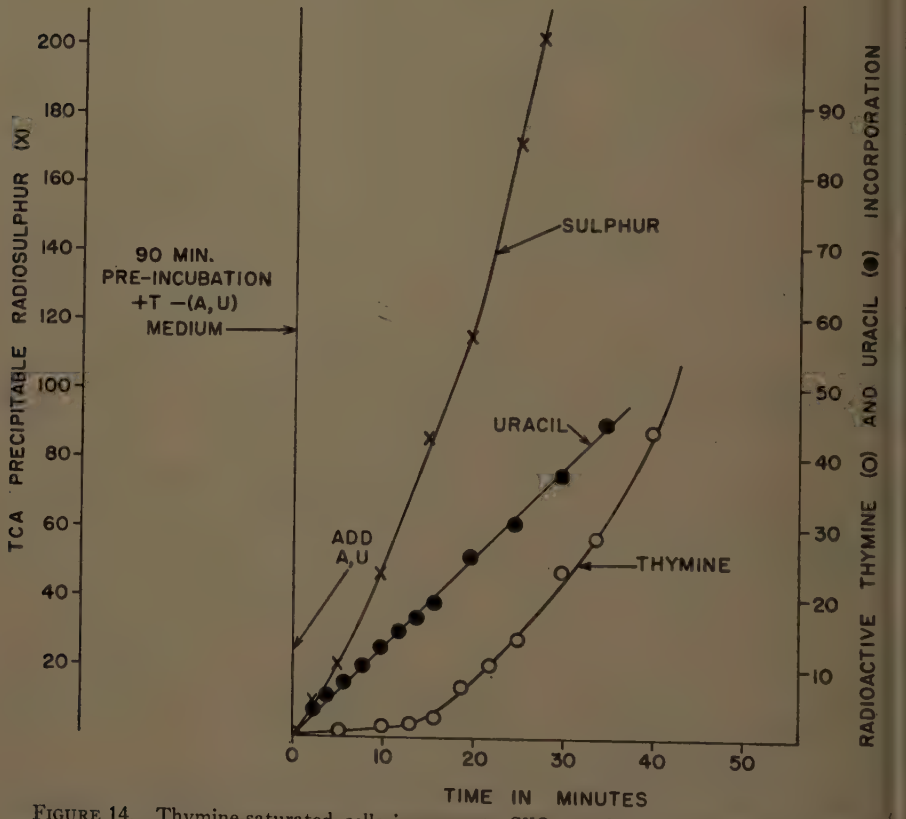


FIGURE 14. Thymine-saturated cells incorporate $S^{35}O_4^{=}$ and C^{14} uracil promptly when arginine and uracil are restored to the medium. C^{14} thymine incorporation begins after a 12-min. lag.

We have tried to find reasons other than nuclear changes that might account for the lag in the inducibility of these cells. Energy and material for synthesis are obviously available since other proteins and ribosomes are made. A delay in the entry of the inducer does not seem to be responsible since the lag is not reduced by increasing the concentration of the inducer one hundredfold. DNA *synthesis* is not required as the enzyme induction proceeds at the same rate whether or not thymine is present. The enzyme-forming system seems to be present and active as there is no lag in the synthesis of the enzyme at the uninduced rate. The lag in both enzyme induction and in thymine incorporation persists after 20 min. incubation with arginine and fluorouracil.

From these results, it appears that the induction of this enzyme is dependent on the state of the DNA of the nucleus. It seems that the same changes (caused by synthesis of RNA and protein) that are required for the resumption

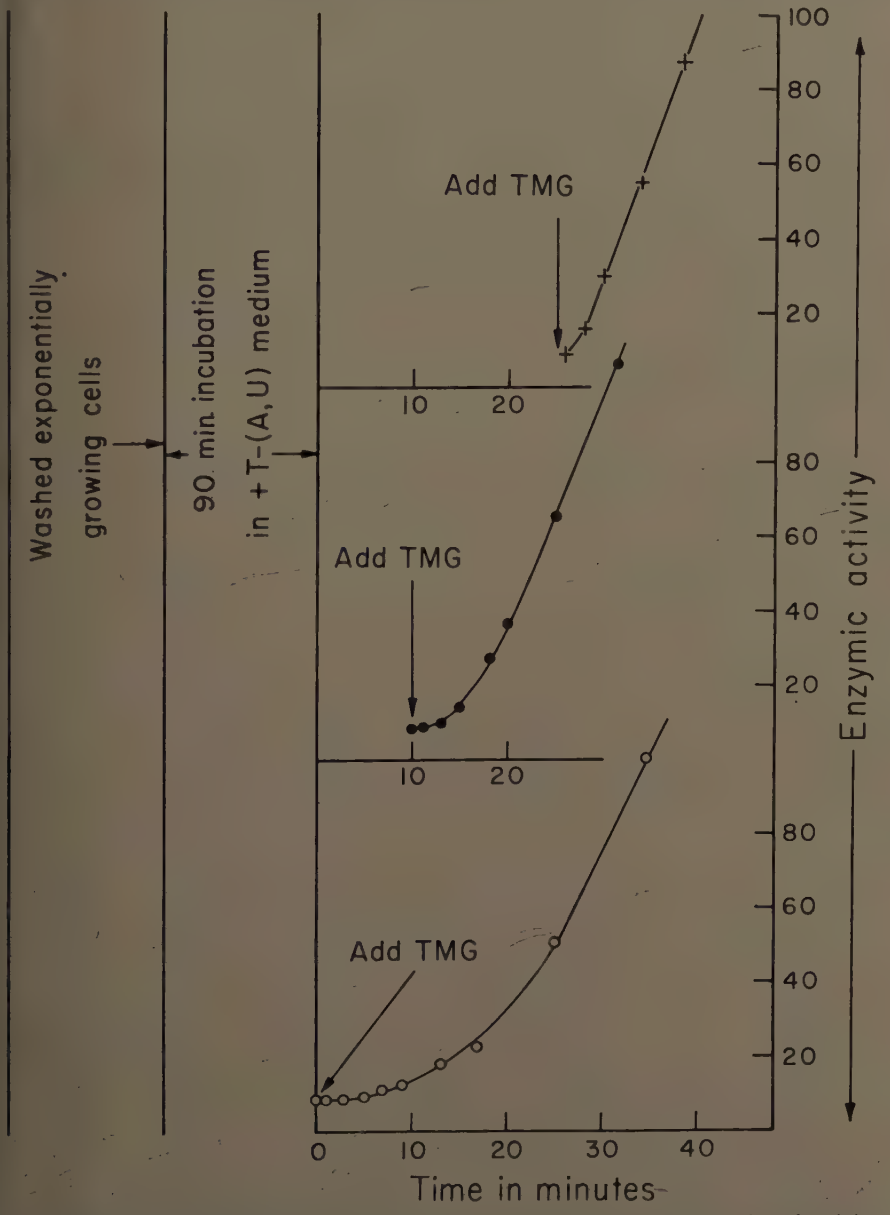


FIGURE 15. Thymine-saturated cells induced at various times after restoration of arginine and uracil. The cells regain capacity to respond to the inducer at the same linear rate whether or not the inducer is present.

of DNA synthesis are also required for the induction process. A tentative interpretation of this relationship is that the inducer acts by facilitating an association between ribosomes and DNA. As a result, the ribosome assumes a configuration more suitable for rapid synthesis. In the thymine-saturated cells neither this association nor synthesis of DNA is possible because the DNA has completed its duplication and is in a four-stranded configuration. After a period of RNA and protein synthesis, nuclear division occurs and DNA synthesis and enzyme induction can be initiated.

This hypothesis is compatible with all of our data to date. It gains support from the work of Nisman[20] who has recently reported stimulation of β-galactosidase synthesis by DNA in a cell-free system. To carry speculation even further, this hypothesis suggests a possible explanation for the correspondence between the order of the genes and the order of biochemical events. It also suggests that the rapid (and inducible) synthesis of protein observed in bacteria is carried out by ribosomes associated with the nucleus, whereas the slow (and noninducible) synthesis of protein in animal cells is accomplished by cytoplasmic ribosomes.

Conclusion

Almost every part of the cell is suspected of playing some role in protein synthesis. Activating enzymes, soluble RNA, lipids, and ribosomes have been under investigation for some time. In some systems, the cell membrane appears to be involved.[21] The present work shows a possible role of DNA. Furthermore, protein synthesis can be halted by a minor disturbance such as an osmotic shock[22] and, therefore, it seems to be quite sensitive to the spatial arrangement of the cellular structures. Disruption of the cell usually decreases the rate of protein synthesis by a factor of a thousand or more. The organization of the synthesizing system appears to be of the greatest importance.

It therefore seems necessary to complement studies of purified systems with studies of intact cells. Otherwise it will be difficult to relate the residual traces of protein synthesis that are observed with cell fragments to the process that occurs in a living organism.

References

1. Roberts, R. B. 1957. Studies of biosynthesis in *Escherichia coli*. Carnegie Inst. Wash. Publ. 607. : 418–442.
2. Bolton, E. T. & R. B. Roberts. 1956. Carnegie Inst. Wash. Year Book No. 55. : 130.
3. Hendler, R. W. 1958. Science. **128:** 143.
4. Britten, R. J. & F. T. McClure. 1958. Carnegie Inst. Wash. Year Book No. 57. : 149.
5. Britten, R. J. & B. J. McCarthy. 1959. Carnegie Inst. Wash. Year Book No. 58. : 266.
6. Bolton, E. T. 1958. Carnegie Institution of Washington Year Book No. 57. : 135.
7. Bolton, E. T., B. H. Hoyer & D. B. Ritter. 1958. *In* Microsomal Particles and Protein Synthesis. : 18–21. Pergamon Press. New York, N. Y.
8. McCarthy, B. J. Biochim. et Biophys. Acta. In press.
9. Britten, R. J. & R. B. Roberts. 1960. Science. **131:** 32–33.
10. Roberts, R. B., R. J. Britten & E. T. Bolton. 1958. *In* Microsomal Particles and Protein Synthesis. : 84–94. Pergamon Press. New York, N. Y.
11. Bolton, E. T. 1959. Carnegie Inst. Wash. Year Book No. 58. : 274.
12. McQuillen, K., R. B. Roberts & R. J. Britten. 1959. Proc. Nat. Acad. Sci. U. S. **45:** 1437–1447.

13. Cowie, D. B., S. Spiegelman & R. B. Roberts. In preparation.
14. Duerksen, J. D. & D. B. Cowie. 1960. Abstr., Fourth Annual Meeting of the Biophysical Society.
15. Britten, R. J., E. T. Bolton, K. McQuillen & R. B. Roberts. 1959. Carnegie Inst. Wash. Year Book No. 58. : 289.
16. McCarthy, B. J. 1959. Carnegie Inst. Wash. Year Book No. 58. : 271.
17. Caro, L. G. & F. Forro, Jr. 1960. Abstr., Fourth Annual Meeting of the Biophysical Society.
18. Barner, H. D. & S. S. Cohen. 1954. J. Bacteriol. **68:** 80.
19. Cowie, D. B., B. J. McCarthy & R. B. Roberts. 1960. Abstr., Fourth Annual Meeting of the Biophysical Society.
20. Nisman, B. & H. Fukuhara. 1960. Compt. rend. sci. **250:** 410–412.
21. Spiegelman, S. 1958. Proc. Intern. Congr. Microbiol. (7th Congr.) Stockholm, Sweden.
22. Roberts, R. B. 1957. Carnegie Inst. Wash. Year Book No. 56. : 124.

Printed in the United States of America

MONOGRAPHIC PUBLICATIONS

OF

THE NEW YORK ACADEMY OF SCIENCES

(Lyceum of Natural History, 1817–1876)

(1) The Annals (octavo series), established in 1823, contain the scientific contributions and reports of researches, together with the records of meetings of the Academy. The articles that comprise each volume are printed separately, each in its own cover, and are distributed immediately upon publication. The prices of the separate articles depend upon their length and the number of illustrations, and may be ascertained upon application to the Executive Director of the Academy.

Current numbers of the Annals are sent free to all Members of the Academy desiring them.

(2) The Special Publications, established in 1939, are issued at irregular intervals as clothbound volumes. The price of each volume will be advertised at time of issue.

(3) The Memoirs (quarto series), established in 1895, are issued at irregular intervals. It is intended that each volume shall be devoted to monographs relating to some particular department of science. Volume I, Part 1 is devoted to Astronomical Memoirs, Volume II to Zoological Memoirs. No more parts of the Memoirs have been published to date. The price is one dollar per part.

(4) The Scientific Survey of Porto Rico and the Virgin Islands (octavo series), established in 1919, gives the detailed reports of the anthropological, botanical, geological, paleontological, zoological, and meteorological surveys of these islands.

Subscriptions and inquiries concerning current and back numbers of any of the publications of the Academy should be addressed to

EXECUTIVE DIRECTOR
The New York Academy of Sciences
2 East Sixty-third Street
New York 21, N. Y.